The molecular evolution of planktic foraminifera and its implications for the fossil record

Iain A. Stewart

Doctor of Philosophy

University of Edinburgh

February 2000

Preface and Declaration

Since graduating from The University of Dundee in 1996 with a MA (Joint Hons) degree in Environmental Science and Geography, the author has been engaged in a programme of full time research under the supervision of Dr. Kate Darling and Prof. Dick Kroon at The University of Edinburgh, Department of Geology and Geophysics.

This thesis is my own work, unless otherwise cited. No part of the work referred to in this thesis has been submitted in support of an application for another degree or qualification from this or any other university or other institute of learning.

Abstract

The marine microfossils of planktic foraminifers are widely used for investigating palaeoceanographic and palaeoclimatic conditions. The objective of this project was to investigate genotypic variation within planktic foraminiferal morphospecies and the spatial distribution of genotypes in the subpolar, transitional and subtropical North Atlantic. Foraminiferal genomic DNA was extracted and the ~1000 base pair 3' terminal region of the small subunit ribosomal RNA gene was amplified using the polymerase chain reaction. Using distance-based molecular phylogenetic analysis, a neighbour-joining phylogeny was reconstructed based on 31 planktic and 15 benthic previously sequenced foraminifera and extended to include 15 genotype sequences obtained from the North Atlantic during this study. Bulk plankton samples were collected for preliminary examination of genotype/morphotype relationships.

The molecular phylogeny is largely consistent with the foraminiferal fossil record. It supports the suggestion that the origins of planktic foraminifers are polyphyletic, as the spinose planktic foraminifers cluster separately from the non-spinose planktic foraminifers within the phylogeny. Branch length variation within the planktic cluster reflects large differences in evolution rate between morphospecies. Within the North Atlantic, genotypic variation has been identified within the morphospecies, *Globigerina bulloides, Turborotalita quinqueloba, Globigerinella siphonifera, Globigerinella calida, Globigerinoides ruber* and *Neogloboquadrina pachyderma*. The distribution of genotypes is complex, and it has been found that genotypes, representing a single morphospecies, often co-exist within the water column. This could be indicative of cryptic speciation, suggesting that North Atlantic planktic foraminiferal diversity is much higher than fossil record interpretations have indicated. The genotypes within *G. bulloides*, *G. siphonifera, G. calida* and *T. quinqueloba* have different geographic distributions within the North Atlantic. It is apparent that *G. bulloides* Types IIa and IIb and *G. siphonifera* Types IIa and IIb have extensive distributions suggesting that they are more generalist in adaptation, and tolerant to a wide range of oceanic conditions. In contrast, *G.*

bulloides Type Ib and possibly *G. siphonifera* Type I may be cold intolerant, indicating that they are likely to be warmer water specialists.

Mapping genotype to morphotype has been made difficult due to the genotypes within a morphospecies living together in the water column, and also due the low number of mature specimens obtained during bulk plankton collection. However, where genotypes have been found as sole representatives of a morphospecies within a region, the bulk sample morphology will reflect the variation within the genotype. In addition, *Globigerina falconensis* was represented by a single genotype within the transitional-subtropical region of the North Atlantic, so may prove more suitable than *G. bulloides* as a transitional-subtropical palaeoceanographic proxy.

Comparison of the North Atlantic genotypes with those from the Southern and Pacific Oceans shows that in some genotypes mixing is sufficient to maintain genetic homogeneity, indicating recent gene flow. The extent to which genotypic variation exists within North Atlantic planktic foraminiferal morphospecies has significant implications for palaeoceanographic studies. Genotypes may be adapted to different environments and if pooled they may provide potentially misleading palaeoceanographic proxies. However, if the genotypes were identifiable using morphological, chemical or isotopic parameters they could provide new high-resolution proxies for palaeocean/climate reconstruction.

Acknowledgements

I would like to thank my supervisors Dr. Kate Darling and Prof. Dick Kroon for all the help, support and encouragement they have provided throughout my PhD. In addition, I would also thank Dr. Chris Wade (the 3rd "unofficial supervisor") for his help and support over the last few years.

I am very grateful for all the support I have received from everybody at The Waddington Building (past and present), particularly Dr. Dave Yirrell, Dr. Marlynne Quigg, Professor Andy Leigh-Brown, Dr. Sarah Lockett and Heather Precious. I would also like to thank those from the Department of Geology and Geophysics who have made this an enjoyable three years. I would like to thank Ruth Acheson for her helpful discussions and to Dr. John Craven for his help with the SEM.

I would like to thank Dr. Simon Troelstra (Department of Earthsciences, Free University, Amsterdam, Netherlands), Dr. Thomas J. Müller (Institute of Meereskunde, Kiel, Germany) and Dr. Susanne Neuer (University of Bremen, Germany) for enabling participation of the research cruises. Also Dr. Octavio Llinas (Marine Research Institute, Telde, Gran Canaria) for help during my visit to his research institute.

The Denmark Strait Expedition was financed by the Netherlands Foundation for Scientific Research (NWO) with additional support by the Geological Survey of Denmark and Greenland (GEUS). I would like to thank the master and crew of RV Professor Logachev, and Dr. Antoon Kuijpers and colleagues for their support during the expedition. Thanks to the master and crew of FS Meteor (M37/2, M38/2) and FS Poseidon (P247), and colleagues for their support during these expeditions. I would also like to thank NERC for funding this project.

A big thanks to my mum and dad for all their support and encouragement over the years. Don't worry I think I'll stop being a student and get a job ! And last but certainly not least; Lucy, thanks for getting me through these last few years.....without you this definitely would have been possible, or worth it !!

Contents

Abbreviations

A	Adenine
AAC	Antarctic Circumpolar Current
ABI	Applied Biosystems
AC	Azores Current
AF	Azores Front
APS	Ammonium persulphate
bp	Base pair
°C	degrees centigrade
C	Cytosine
CC	Canary Current
CLIMAP	Climate: Long-Range Investigation, Mapping and Prediction
CTAB	Cetyltrimethylammonium bromide
D	Dextral (right) coiling
depc	Diethylpyrocarbonate
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleoside triphosphates
EDTA	Ethylenediaminetetra-acetic acid
EGC	East Greenland Current
ESTOC	European Station for Time Series in the Ocean, Canary Islands
et al.	*et alli* (and others)
FE	Formamide ethylenediaminetetra-acetic acid
G	Guanine
GDE	Genetic Data Environment
GEUS	Geological Survey of Denmark and Greenland
HCl	Hydrochloric acid
IC	Irminger Current
kg	Kilogram
LMP	Low melting point

LSU	Large subunit
M37/2	Meteor Cruise. Canary Islands (28/12/96 to 5/1/1997)
M37/2	Meteor Cruise. Canary Islands (12/4/97 to 14/4/97)
Ma	Millions years before present
$MgCl_2$	Magnesium chloride
µg	microgram
µl	microlitre
µm	micron
NAC	North Atlantic Current
NaCl	Sodium chloride
NaOCl	Sodium perchlorate
NATW	North Atlantic Transitional Water
nm	nanometer
NERC	Natural Environment Research Council
NWO	Netherlands Foundation for Scientific Research
pm	picomolar
PCR	Polymerase Chain Reaction
PHYLIP	Phylogenetic Inference Package
r	Ribosomal
RAPD	Randomly Amplified Polymorphic DNA
RNA	Ribonucleic acid
S	Sinistral (left) coiling
SDS	Lauryl sulphate
SEM	Scanning electron microscopy
SST	Sea surface temperature
SSU	Small subunit
T	Thiamine
Taq	*Thermus aquaticus*
TBE	Tris/borate EDTA
TEMED	N, N, N', N'- Tetramethylethylenediamine
Tris	Hydroxymethyl-aminomethane

tRNA	Transfer RNA
U.V.	Ultra violet
v	volt
w/v	weight to volume

Chapter 1: Introduction Index

1.1. Introduction to planktic foraminifera

Foraminifers are single celled (unicellular) eukaryotes[1] from the Kingdom Protista. Most foraminiferal groups have a calcitic shell (test) and their fossil record extends back approximately 550 Ma (Culver, 1991). The foraminifera can be divided up into two main groups based on their life strategy. One group are the benthic foraminifera which live in or on the sediments in the deep-ocean, in or on near-shore sediments and also on marine vegetation, and the other group are the planktonic (planktic) foraminifera which inhabit the upper few hundred metres of the water column within the world oceans. The planktic foraminifera form a significant constituent part of the marine zooplankton from polar to tropical water masses (Bé, 1977). The fossil record indicates that the first planktic foraminiferal groups evolved from the benthic group during the early Jurassic, approximately 200 million years ago (Görög, 1994).

After reproduction or death, the planktic foraminiferal tests sink down through the water column to the ocean floor. The calcite tests form an important component of the biogenic particle flux within the oceans and upon settling on the ocean floor, the fossil foraminifers constitute a substantial part of the deep-sea marine sediments. The fossil foraminifers are preserved within the marine sediments and their fossil record has provided one of the most extensively used tools with which to examine palaeoceanographic and palaeoclimatic conditions since the mid-Jurassic. In addition, the planktic foraminifera have also provided a tool for stratigraphic correlation between deep-sea sediment cores, and for the

[1] Eukaryotes can be defined as individuals which possess a membrane-bound nucleus and chromosomes, and extensive intra-cellular compartmentalisation.

geochronological dating of the marine sediments (e.g. Deep-Sea Drilling Project and the Ocean Drilling Project Initial Reports, 1969-present).

In the present day, there are 45 extant morphologically-defined planktic species (morphospecies) described, of which 21 morphospecies commonly occur within the world's oceans (Hemleben *et al.*, 1989). Some of the original descriptions of the modern planktic foraminifera date back to the 19th century (e.g. Brady, 1879; d'Orbigny, 1826 and Ehrenberg, 1861), with further extant morphospecies being described during the 20th century (e.g. Cifelli, 1961; Natland, 1938 and Parker, 1962). There have been many taxonomic revisions over the years, as planktic foraminiferal research has grown and developed, accompanied by often heated taxonomic debate.

Since the late 1950's, there has been an enormous volume of work produced about planktic foraminifers. Much of the earlier research was involved with distributional studies of the various morphospecies and investigation of morphological variability, together with taxonomic debate. As research has developed, aided by technological improvements, utilisation of the planktic foraminifera as a palaeoceanographic tool has proved extremely valuable.

1.2. Distribution of extant planktic foraminifera

There have been a number of key research papers describing the distribution of extant planktic foraminifers within the Atlantic, Pacific and Indian Oceans, based on sediments obtained from the ocean floor and also from surface plankton net tows (Barrett Smith, 1963; Bé and Hamlin, 1967; Bé and Tolderlund, 1971; Boltovskoy, 1971; Bradshaw, 1959; Kennett, 1968, Parker, 1962; Tolderlund and Bé, 1971).

These enabled Bé (1977) to classify Recent planktic foraminifera into five faunal assemblage zones: tropical, subtropical, transitional, subpolar and polar provinces (Fig. 1-1). These provinces are mirror imaged in each hemisphere reflecting similar oceanic conditions.

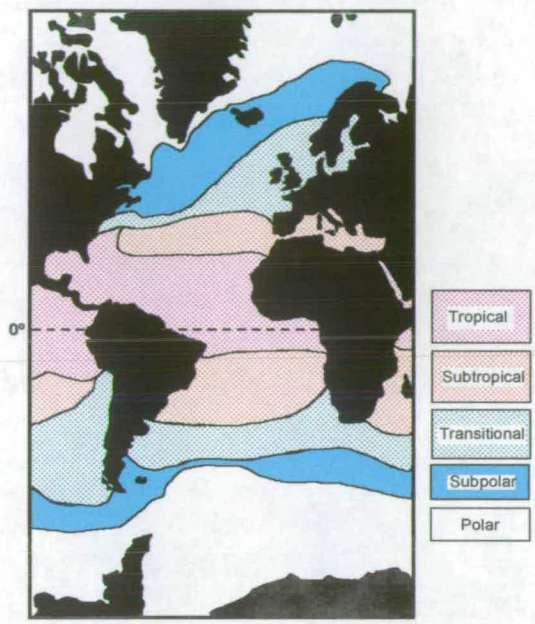


Fig. 1-1. Planktic foraminifer faunal provinces of the Atlantic Ocean (modified from Bé, 1977).

However, it has long been understood that large scale general distributional studies can only provide part of the picture, since the actual distribution in the ocean is clearly highly complex with seasonal effects resulting in spatial and temporal variations in planktic foraminiferal abundances (Bradshaw, 1959; Bé, 1960; Tolderlund and Bé, 1971). Recent sediment trap data has provided high resolution records of morphospecies composition and relative abundance within the water column to address these issues (e.g. Reynolds and Thunell, 1986; Deuser *et al.*, 1981; Deuser, 1987; Deuser and Ross, 1989; Sautter and Thunell, 1991). The sediment trap data provides important information regarding the seasonal succession of planktic

foraminiferal morphospecies, which can be related to the prevailing hydrography at the collection location. This combined with variations in species abundance and isotopic compositions provide detailed ecological information of use in palaeoceanographic investigations.

1.3. The use of planktic foraminifera in palaeoceanography

The marine microfossils of planktic foraminifers are widely used for reconstructing palaeoenvironments, such as climate, ocean circulation and primary productivity. There are a number of ways in which this can be achieved; namely through the use of morphological variation, absolute and relative abundances of morphospecies, transfer functions, coiling direction, and stable isotopes.

1.3.1. Morphological variation

There is an inherent limitation to the use of planktic foraminifers as proxies of climate change since defining morphospecies based on morphological evidence is entirely subjective. Many taxonomic arguments have arisen largely depending on whether someone is a "lumper" (one who groups many morphotypes under the name of a single morphospecies) or a "splitter" (one who divides the morphotypes up into many separate morphospecies) (Haynes, 1992). It is therefore impossible to be certain whether morphological variants are different species or whether the variation reflects a response to different environmental conditions.

A number of investigations have examined the distributional pattern of within species morphological variation. Kennett (1968) examined *Neogloboquadrina pachyderma* from sediment core-tops of the southwest Pacific Ocean. He found that

there were three major groups: (1) A southern group in the polar Antarctic waters. These were predominantly sinistral coiled, 4 chambered and had a thick test; (2) An intermediate group in the northerly Antarctic waters. These were predominantly sinistral coiled, 4 ½ to 5 chambered and had a thin test; (3) A northern group in the northern subantarctic to subtropical waters. These were dominantly dextral coiled, 4 chambered and had a thin test. The distribution limits of these morphotype groups in the sediment does not quite fit the present day location of ocean frontal systems, since the three groups are ~8° further north than the Antarctic fronts (Kennett, 1968). A later study of sediment core-tops in the same area, by Malmgren and Kennett (1972), showed that the frequency and coiling direction of *N. pachyderma* specimens was related to latitude.

The distribution of within species morphological variation has also been examined in *Orbulina universa*. Using sediment core tops from the Indian Ocean, Bé *et al.* (1976) found that test diameter and test porosity in *O. universa* varied with latitude. Specimens found within the tropical region had a larger test diameter and greater test porosity than those found within the subtropical region. Subsequently, combining data from sediment core tops and plankton tows, Hecht *et al.* (1976) investigated *O. universa* morphological provinces in the Indian Ocean. They found a major discontinuity in the gradient of test porosity between the equatorial and central Indian Ocean, indicating the presence of two morphological groups. The division between the two populations corresponds approximately to an oceanic front located at ~10°S. Hecht *et al.* (1976) suggested that the geographic variation of test porosity may, therefore, reflect ecological differences between the populations from the central and equatorial Indian Ocean. Further, comparison of open-ocean *versus*

ocean-margin *O. universa* assemblages, by Robbins (1988), revealed that abnormal morphotypes were more frequent in ocean-margin samples, and that this was correlated to high primary productivity.

In many cases it has been suggested that the morphological variation found within a species may be a phenotypic effect induced by changes in ocean environment characteristics, such as temperature, salinity or productivity. A number of laboratory culture experiments have attempted to determine the effects of temperature and salinity on the morphology of foraminiferal specimens (e.g. Hemleben *et al.*, 1987; Bijma *et al.*, 1990a). However, Hemleben *et al.* (1987) could not establish a link between environmental parameters and the pre-gametogenic morphology of *Globigerinella sacculifer*. The investigation by Bijma *et al.* (1990a) of seven species of planktic foraminifera showed that, in general, optimum temperature and salinity produced the largest mean final test sizes. In addition, it was also noted that, contrary to previous investigations, the frequency of "abnormal" phenotypes was higher during normal conditions, where as the frequency of "normal" phenotypes was higher during extreme conditions (Bijma *et al.*, 1990a).

In addition to the studies outlined above, *T. quinqueloba* specimens from the North Atlantic-Mediterranean were noted to differ morphologically from northern Indian Ocean specimens (Kroon *et al.*, 1988). They suggested two explanations: (1) the morphological differences were either the result of temperature-induced phenotypic variation or (2) the Atlantic-Mediterranean and the northern Indian Ocean morphotypes are separate subspecies or genotypes. Further, Brummer and Kroon (1988) examined coiling direction ratios of planktic foraminifers in the Atlantic and Indian Oceans. Since the non-spinose planktic foraminifers displayed

provinces of either left (sinistral) or right (dextral) coiling, they concluded that coiling direction in non-spinose morphospecies was a genetically controlled binary trait, rather than a phenotypic effect induced by the ocean environment.

Perhaps one of the most contentious debates in foraminiferal research is whether the morphological variation observed within a species is a result of natural variation within a population, environment induced phenotypic variation, or whether they represent separate species (Darling *et al.*, 1999). This, of course, has considerable implications for palaeoceanographic investigations utilising planktic foraminifera.

1.3.2. Morphospecies abundances

The absolute and relative abundance of each morphospecies within a plankton assemblage collected along a transect can be correlated with the corresponding hydrographic regime (Ufkes *et al.*, 1998; Ottens, 1991, 1992). For instance, a plankton sample from a tropical or subtropical region containing predominantly cool water morphospecies such as *G. bulloides* is indicative of upwelling (Kroon, 1991; Kroon and Ganssen, 1989). Such relationships between the planktic foraminiferal assemblage and hydrographic regime are reflected in the sedimentary record, and can therefore be used to indicate similar palaeocean environments. For example, a high abundance of *G. bulloides* within a low-latitude fossil assemblage is considered an upwelling indicator (Naidu and Malmgren, 1996b). Similarly, the relative abundances of foraminiferal morphospecies within the marine sediments are also used in analyses utilising transfer functions.

1.3.3. Transfer Functions

The use of transfer functions has provided paleoceanographers with a valuable source of information for estimating past SSTs. There are basically two different methods for achieving this. The first is the method developed by Imbrie and Kipp (1971) which has been successfully used by CLIMAP (1984) (Climate: Long-Range Investigation, Mapping and Prediction) to reconstruct palaeo-sea surface temperatures (SSTs) for the Last Glacial Maximum and the Last Interglacial Climatic Optimum. This method can suffer from a no-analogue situation, whereby the present-day fauna is so different from the fossil fauna that it cannot fit into the model, such that the resolution of the data is reduced. The second method is the modern analogue technique, based on the level of dissimilarity between modern and fossil assemblages, and has been used quite extensively (e.g. McIntyre *et al.*, 1989; Pflaumann *et al.*, 1996). Both of these techniques have a number of inherent assumptions. (1) The ecology of the species assemblages has not changed since the time of the estimated SST, (2) The faunal assemblage found in the sediments is systematically related to the SST, (3) The caloric winter is colder than the caloric summer, even in tropical areas under the influence of upwelling, (4) The "winter" and "summer" environment of the northern and southern hemisphere is not significantly different (after Pflaumann *et al.*, 1996, p15). A subjective element is introduced into these investigations with the selection of the most suitable analogues for analysis, and also the taxonomic interpretation of the foraminiferal morphospecies.

1.3.4. Planktic foraminifer coiling direction

Micropaleontologists have used the coiling direction of specific planktic foraminiferal morphospecies for different purposes. Initial interest in coiling direction was fuelled by the observation of sinistral and dextral coiling provinces, which could be used to correlate stratigraphic zones in deep-sea sediment cores (Ericson *et al.*, 1954; Ericson and Wollin, 1956; Saito, 1976). The importance of coiling direction ratios within *Neogloboquadrina pachyderma* was recognised (Ericson, 1959; Bandy, 1960; Bandy, 1972), and fossil assemblages dominated by sinistral coiling *Neogloboquadrina pachyderma* were used to indicate the presence of specific ocean isotherms. More recently, the enhanced flux of sinistral coiling forms of *G. bulloides* and *N. pachyderma* within the fossil record have been used to indicate periods of intensified upwelling (e.g. Naidu and Malmgren, 1996a).

1.3.5. Geochemical proxies

The stable isotopes, $\delta^{18}O$ and $\delta^{13}C$, incorporated within the calcite of foraminiferal tests have provided an important source of geochemical information for palaeoceanographic investigation (Epstein *et al.*, 1953; Bard *et al.*, 1987; Bemis *et al.*, 1998; Ganssen and Kroon, in press; Sautter and Thunell, 1991b; Marchitto Jr. *et al.*, 1998). The ratio of ^{16}O to ^{18}O within the test calcite is mainly temperature dependent, with a higher proportion of ^{18}O reflecting a cooler (and/or deeper) habitat. Thus, the sedimentary $\delta^{18}O$ record has provided an extensive record of the changes in palaeo-SSTs, such as with the glacial/interglacial cycling of the Quaternary. The $^{12}C/^{13}C$ ratio within the water column is affected by photosynthetic activity. Photosynthetic algae preferentially utilise ^{12}C and the surface waters tend to become enriched in ^{13}C.

Therefore, a higher proportion of ^{13}C within test calcite reflects a shallower habitat (Bijma *et al.*, 1998). The meaningfulness of $^{12}C/^{13}C$ ratio is complex and its use as proxy attracts many scientific debates.

Recently, foraminiferal research has also utilised Cd/Ca and Mg/Ca ratios as estimates of palaeoproductivity and palaeotemperature (e.g. Elderfield *et al.*, 1998; Rickaby and Elderfield, 1999).

1.4. The development of biochemical and genetic studies for foraminiferal research

The first attempts at using biochemical data for the investigation of planktic foraminifera were made by King and Hare (1972a,b). Targeting sixteen planktic morphospecies from deep-sea sediment core-tops, they examined the composition of amino acids preserved in the proteinaceous organic layer of the foraminiferal tests. The authors suggested that the data might prove useful for resolving species, subspecies or phenotypic variants. However, the actual information obtained was of limited use. More recently, Robbins and Healy-Williams (1991) attempted to link stable isotope analysis, morphological data and biochemical information. However, again, since amino acids from the proteinaceous organic layer of the test were analysed, the biochemical information obtained was of limited use.

Unfortunately, neither DNA nor RNA are found in the proteinaceous organic layers of foraminiferal tests and are therefore not preserved in the fossil record. Genetic investigations of foraminifera have therefore turned to examining the genome of extant species. This has led to the utilisation of molecular phylogenetics

to provide an insight into the evolutionary relationships within foraminifera. Up to the present, the investigations have generally been centred on the use of the ribosomal (r) RNA gene which is providing new and exciting information about foraminiferal evolution and between taxa relationships.

1.4.1. The form and function of the rRNA gene

Ribosomes are complex organelles found within the cytoplasm of all eukaryotic cells, which interpret the genetic code contained within nuclear DNA and translate the information to synthesise proteins. The ribosomal RNA expressed by the rRNA gene forms an essential component of the 3-dimensional structure of the ribosome. The constancy of the structural form of the ribosome is crucial to the successful translation of the genetic code. Therefore, the rRNA genes have conservation mechanisms to maintain their structural integrity. The ribosomal array consists of a series of tandem repeats of the unit shown in Fig. 1-2. The number of repeats within an individual genome can vary from a single copy to possibly many hundreds of copies (Page and Holmes, 1998). The multiple copies permit ribosomes to be formed and proteins to be synthesised rapidly whenever required by the cell.

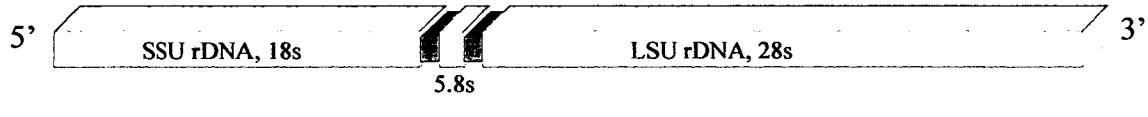


Fig. 1-2. The rRNA array (after Hillis and Dixon, 1991). The small subunit (SSU), large subunit (LSU), and the 5.8s RNA molecules are shown. The blue regions indicate the internal transcribe spacers, which separate the RNA genes.

Early molecular research into the evolution and phylogenetic relationship of prokaryotic and eukaryotic organisms concentrated on the use of ribosomal (r) DNA

nucleotide sequences. This has provided extensive databases of small subunit (SSU) and large subunit (LSU) rDNA sequences for use in molecular phylogenetic inference (Neefs *et al.*, 1993; De Rijk *et al.*, 1996; Maidak *et al.*, 1994, 1996). Such a substantial database has enabled an evolutionary "tree of life" to be constructed from the homologous sections of the SSU rRNA gene using representatives from all the domains of cellular life (Olsen and Woese, 1993). In evolutionary biology 'homologous' refers to having a common descent. The rDNA sequences contain both highly conserved and variable regions which make it suitable for reconstructing molecular phylogenetic relationships between both close and distant organisms (Page and Holmes, 1998).

Darling *et al.* (1997) aligned an ~1000 base pair (bp) region of the 3' terminal region of the planktic foraminiferal SSU rRNA gene against a wide range of eukaryotic taxa. It was found that the amplified region consisted of eight conserved regions (C1-C8) separated by seven variable regions (Fig. 1-3). Four of the variable regions (V7a-V9) are common to the other aligned eukaryotic sequences, and three of the variable regions are insertions specific to foraminifera (F1-F3). The arrangement of the amplified fragment is illustrated in Fig. 1-3.

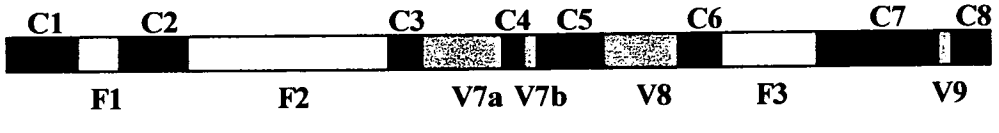


Fig. 1-3. Schematic representation of the ~1000 bp 3' terminal region of the SSU rRNA gene. C1-C8 represent the highly conserved regions which were aligned relative to comparable regions present in all eukaryotes, V7-V9 represent variable length expansion segments present in most eukaryotes and F1-F3 represent three insertions which are unique to foraminifera (after Darling *et al.*, 1997).

1.4.2. The use of molecular data within foraminiferal research

After a series of erroneous attempts to extract and amplify foraminiferal SSU rDNA by Langer *et al.* (1993) and Wray *et al.* (1993), successful amplification was achieved for benthic foraminifera (Pawlowski *et al.*, 1996) and planktic foraminifera (Darling *et al.*, 1996). A number of investigations were also carried out on foraminiferal LSU rDNA (Merle *et al.*, 1994; Pawlowski *et al.*, 1994a,b; Holzmann *et al.*, 1996). The basic problem encountered by Langer *et al.* (1993) and Wray *et al.* (1993) was the isolation of the foraminiferal rDNA from contaminant DNA, such as ingested prey or symbionts (dinoflagellates, diatoms and chrysophytes) that many planktic foraminiferal species possess. Indeed, both studies actually isolated and amplified algal DNA in the belief that it was foraminiferal in origin. The foraminiferal rDNA was correctly isolated by Darling *et al.* (1996b) and Wade *et al.* (1996) gave the following reasons as to why the sequences analysed in this case were foraminiferal:

1. The large insertions (F1-F3, Fig. 1-3) in the SSU rDNA are specific to foraminifera and are not present in any other eukaryotes.

2. Only the foraminiferal rDNA fragments form a distinct monophyletic group within the eukaryote SSU rDNA phylogeny.

3. The phylogenetic placement of the foraminiferal sequences out-with the main crown group radiation (which includes, for example, dinoflagellates and diatoms) is in agreement with the placement based on partial LSU rDNA sequences (Merle *et al.*, 1994; Pawlowski *et al.*, 1994a).

Further statistical analyses by Wade *et al.* (1996) showed that the foraminiferal lineage originates prior to the radiation of the crown group organisms (Fig. 1-4).

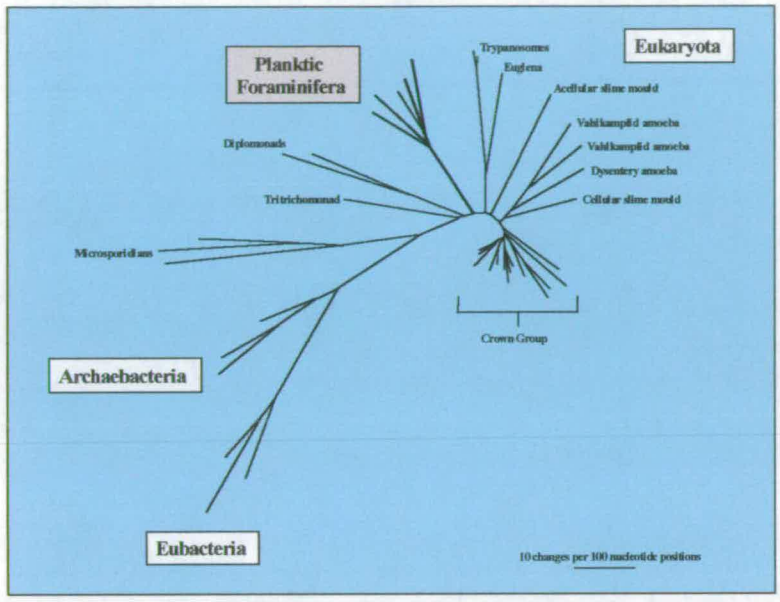


Fig. 1-4. Molecular phylogenetic tree of life (after Wade *et al.*, 1996). The foraminiferal lineage originates prior to the radiation of the crown group organisms. The radiation of the crown group, which includes many of the symbionts and prey of foraminifers, is also shown.

The molecular phylogenetic information from the SSU rRNA gene has enabled investigation of the relationships between foraminiferal morphospecies (Darling *et al.*, 1996b, 1997; de Vargas *et al.*, 1997; Pawlowski *et al.*, 1997). One of the most important early discoveries following the construction of a foraminiferal SSU rDNA phylogeny was that the planktic foraminifers are polyphyletic in origin, with some groups having evolved from separate benthic ancestors (Darling *et al.*, 1997). The molecular phylogeny (Fig. 1-5) shows that the non-spinose planktic foraminiferal morphospecies, *Neogloboquadrina dutertrei*, clusters amongst the benthic foraminifers, separately from the main spinose planktic cluster. This indicates that the planktic foraminifers have evolved from different benthic

ancestors. The polyphyletic origins of other planktic foraminiferal groups has subsequently been confirmed with the addition of further non-spinose planktic foraminifers into the molecular phylogeny (de Vargas *et al.*, 1997; Darling *et al.*, 1999).

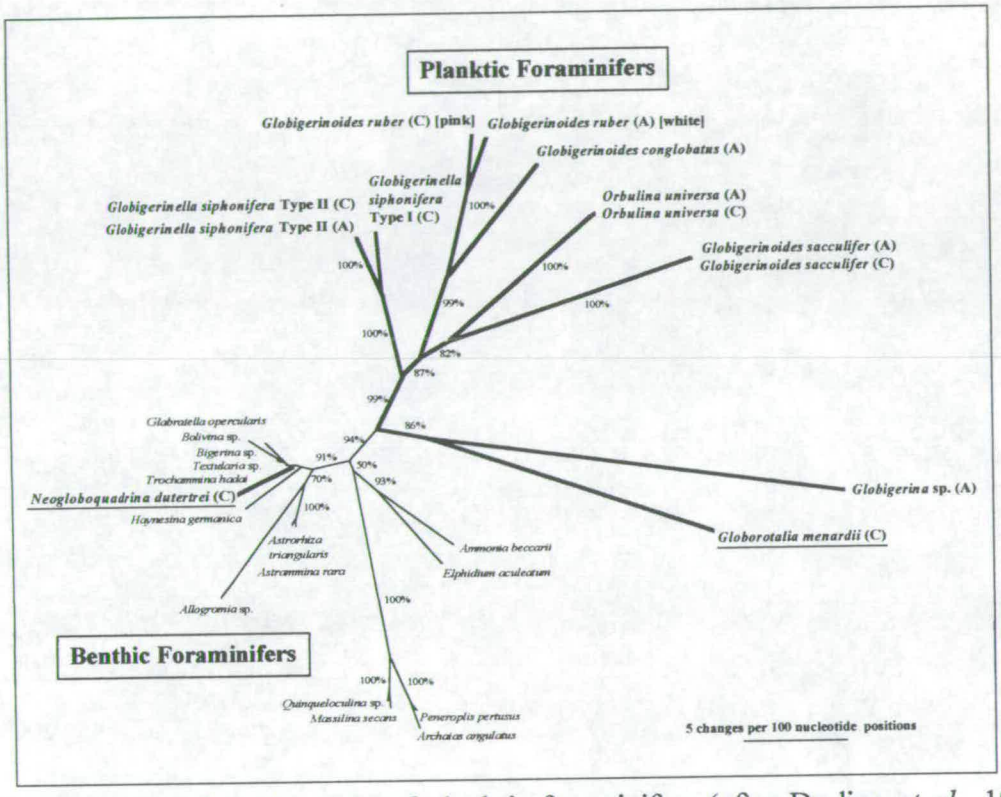


Fig. 1-5. The polyphyletic origin of planktic foraminifera (after Darling *et al.*, 1997). Note that the planktic foraminifer *Neogloboquadrina dutertrei* clusters amongst the benthic foraminifers, separately from the other planktic species (shown in bold).

Comparisons have been made between the molecular phylogenetic data and the fossil record (de Vargas *et al.*, 1997; Darling *et al.*, 1999). It has been observed that the molecular phylogeny is consistent with data from the fossil record, and supports the fossil evidence that the spinose planktic species radiated over a short period of time during the late Oligocene, ~ 30-27 Ma (Darling *et al.*, 1999). Further, it is evident from comparison of molecular phylogenetic analyses to the fossil record that the rate of molecular evolution of the SSU rRNA gene varies considerably

between foraminiferal taxa (Darling *et al.*, 1997, 1999; de Vargas *et al.*, 1997, 1998; Pawlowski *et al.*, 1997). Indeed, Pawlowski *et al.* (1997) estimated that in planktic foraminifera the rate of molecular evolution of the SSU rRNA gene is approximately 50 to 100 times faster than in some benthic foraminifers.

Recently, Darling *et al.* (1999) have examined potential gene flow between planktic foraminiferal populations from different oceans. By comparing genotype representatives of specific morphospecies from tropical, subtropical and transitional faunal provinces, they could determine whether the multi-provincialism of morphospecies reflected genetic isolation. It was discovered that genotypes of *O. universa* and *G. sacculifer* from the Coral Sea (in the Pacific) and the Caribbean Sea were genetically homogeneous (Darling *et al.*, 1999), indicating that genetic mixing has taken place. They suggested that gene mixing in the tropical/subtropical zone most likely occurs from the Pacific to the Atlantic *via* the Cape of South Africa, carried along by the prevailing ocean surface currents.

Recent work has shown that planktic foraminiferal morphospecies are commonly comprised of complexes of genotypes (Darling *et al.*, 1999) indicating the occurrence of previously unrecorded "cryptic" speciation events (Huber *et al.*, 1997; Darling *et al.*, 1999, submitted; de Vargas *et al.*, 1999). Huber *et al.* (1997) completed detailed investigation of two "morphotypes" of *Globigerinella siphonifera*. They used a combination of data sources, which included genetic, morphological, biological and isotopic differences, and concluded that the two morphotypes were distinct cryptic species (see Chapter 5, section 5.1.1). Similarly, de Vargas *et al.* (1999) proposed that the genotypic variation within *Orbulina universa* from the Atlantic also represented cryptic speciation, and demonstrated a

link between genotype distribution and ocean productivity. Further, it has now been shown that the coiling direction of subantarctic *Neogloboquadrina pachyderma* is associated with genetically distinct populations and is not a developmental effect triggered by temperature (Darling *et al.*, submitted).

1.5. The North Atlantic

Planktic foraminifers from the North Atlantic sediments have been widely utilised for palaeoceanographic reconstructions. In particular, much work has been concentrated on sediments of the Quaternary period, which reflect the glacial/interglacial climatic cycling of the Northern Hemisphere. The main hydrographic regions and major surface currents of the Northeastern Atlantic are illustrated in Fig. 1-6. The Northeastern Atlantic is comprised of five water mass regions, including the polar water mass. The regions of importance in this study are the subarctic region, the North Atlantic Current (NAC) waters, the North Atlantic Transitional (NATW) waters and the Azores Current (AC) waters, which are separated by well-defined water mass frontal zones (Fig. 1-6; after Ottens, 1992). The AC and the NATW currents (subtropical) are formed from eastwardly flowing branches of the Gulf Stream. The northeastward flowing NAC (transitional) is an extension of the Gulf Stream. The NAC divides in the mid-latitudes to source the southward flowing Canary Current (CC) (transitional), with the other branch flowing into the higher latitudes. The NAC divides again, and one branch known as the Irminger Current (IC) (transitional/subpolar) flows clockwise round Iceland. Along the east coast of Greenland, the cold East Greenland Current (EGC) (polar) flows southwards over the shelf.

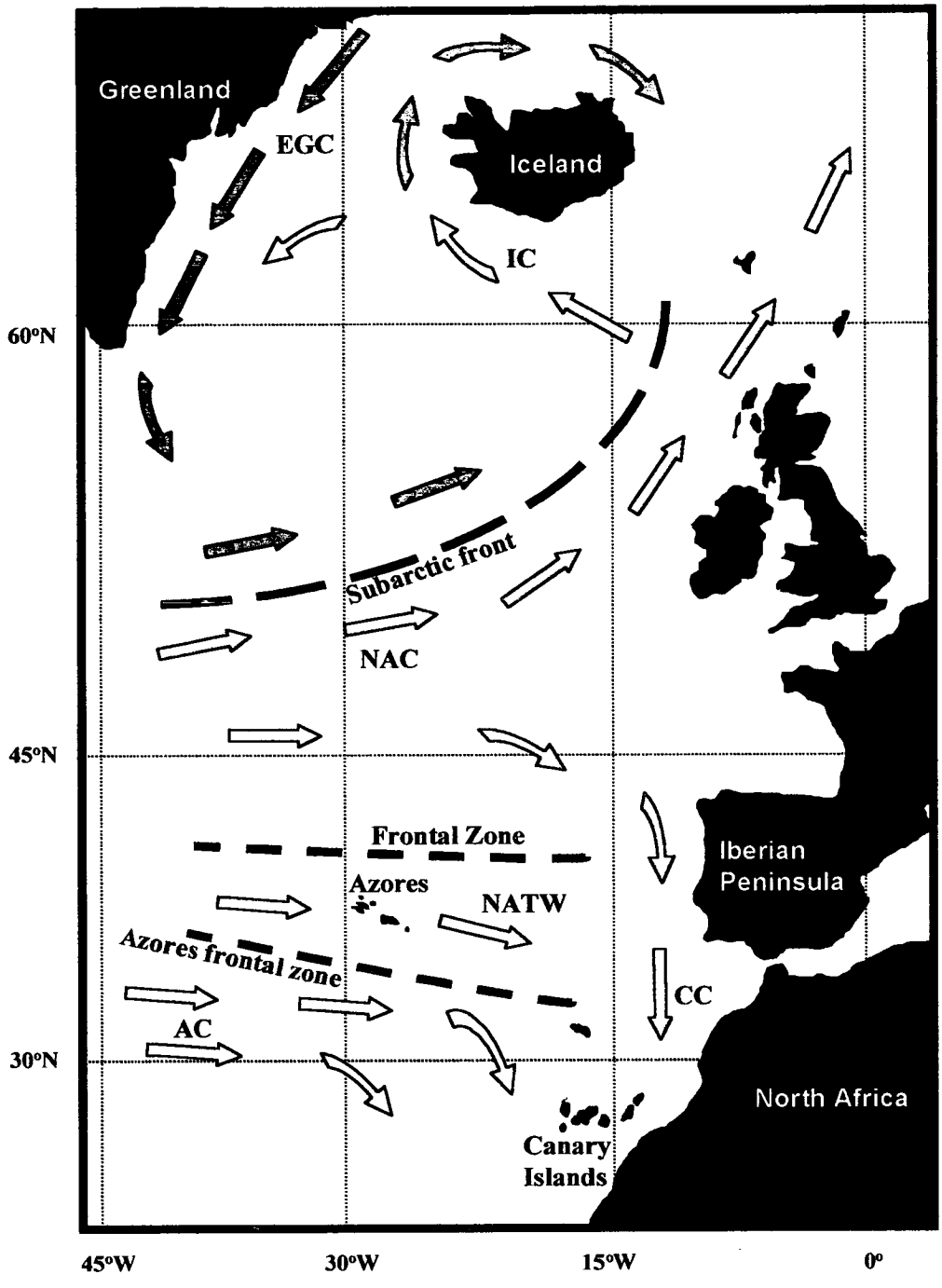


Fig. 1-6. Oceanography of the Northeastern Atlantic (after Ottens, 1992). The position of the frontal zones are denoted by a dashed line. The polar front is not shown due to its proximity to the east Greenland margin. The major surface currents are also shown. The EGC denotes the East Greenland Current, IC the Irminger Current, NAC the North Atlantic Current, AC the Azores Current, CC the Canary Current and NATW the North Atlantic Transitional Water. The surface currents are coloured to represent their relative temperature: blue represents the coldest, and yellow the warmest.

To aid interpretation of the sedimentary record the distribution of the extant planktic foraminifers in the North Atlantic has been widely investigated (e.g. Bé and Hamlin, 1967; Ottens, 1991, 1992). It has now become possible to extend these studies to include molecular phylogenetic investigations of the planktic foraminifers of the Northeastern Atlantic. In this study I have collected planktic foraminifers from subpolar, transitional, subtropical watermasses. Using a molecular phylogenetic approach, a number of paleoceanographic issues can be addressed. The genotypic diversity of planktic foraminifera has been investigated to determine whether North Atlantic morphospecies are composed of multiple genotypes, to identify potential cryptic speciation events. By examining the distribution of genotypes in the North Atlantic, further insight may be gained into whether the genotype distribution can be related to the ocean environment. This may provide additional evidence of cryptic speciation events and will indicate whether genotypes within a morphospecies have different habitat adaptations, which has significant implications for paleoceanographic investigations. The examination of genotypic variation within planktic foraminiferal morphospecies could facilitate the debate regarding the causes of morphological variation. In addition, comparison of genotypes could assist in the understanding of whether genetic isolation or genetic interchange occurs within planktic foraminiferal populations from different regions of the globe. These issues could have significant implications for foraminiferal research.

1.6. Work presented in this thesis

In this study, genotypic variation within North Atlantic subpolar, transitional and subtropical planktic foraminiferal morphospecies has been investigated to

determine whether North Atlantic morphospecies are comprised of multiple genotypes. A total of nine North Atlantic planktic foraminiferal morphospecies have been studied, and the spatial distribution of genotypes from six of these morphospecies has been examined to investigate whether genotypes are distributed randomly or have a specific pattern that may be related to the ocean environment. As foraminiferal tests are destroyed during DNA extraction, bulk plankton samples were collected to investigate possible genotype/morphotype relationships. The molecular data has enabled investigation of the following topics:

1. Molecular phylogenetic relationship between planktic foraminiferal morphospecies.
2. Genotypic variation within planktic foraminiferal morphospecies.
3. Distribution of genotypes and their relationship to the ocean environment.
4. Preliminary investigation of genotype/morphotype relationships.
5. Potential gene flow.

The methodology adopted for both field and laboratory work is described in Chapter 2. Chapter 3 details the reconstructed foraminiferal molecular phylogeny and a comparison is made between the topology of the phylogenetic tree with previous interpretations of the fossil record. Chapters 4, 5, 6 and 7 provide detailed description and discussion of individual monophyletic groups as identified from the molecular phylogeny. They are grouped as follows:

Chapter 4 – *Globigerina/Turborotalita* cluster, comprising *G. bulloides*, *G. falconensis*, and *T. quinqueloba*.

Chapter 5 – *Globigerinella* cluster, comprising *G. siphonifera* and *G. calida*.

Chapter 6 – *Globigerinoides ruber*.

Chapter 7 – Non-spinose morphospecies, comprising *Neogloboquadrina* species and *Globigerinita uvula*.

Finally, the discussion in Chapter 8 is divided into three main sections. Firstly, genotypic variation within North Atlantic planktic foraminiferal morphospecies is examined, followed by an examination of genetic exchange between Atlantic bipolar populations and between North Atlantic and Pacific populations. The final section looks at the broader implications of this study and suggests further work for investigation.

Chapter 2: Materials and methods Index

2.1. Introduction

In this chapter the methodology adopted for both the field work and subsequent laboratory work is described. The first section describes the plankton sampling techniques and localities. The next section is a detailed description of the molecular biology techniques employed during this research, illustrating the methodology used to extract, amplify and sequence the planktic foraminiferal small subunit (SSU) ribosomal (r) DNA. This is followed by an explanation of molecular phylogenetic analysis, including a description of how a molecular phylogeny is constructed, and the preferred methods employed during this study.

2.2. Plankton collection

2.2.1. Canary Island collections onboard FS METEOR

Plankton samples were collected whilst onboard the German ship FS Meteor in the waters around the Canary Islands on cruises M37/2 (28th December 1996 – 5th January 1997) and M38/2 (12-14th April, 1997). The main purpose for both of the cruises was the servicing of a sediment trap at ESTOC (European Station for Time Series in the Ocean, Canary Islands), 100km north of Gran Canaria. The area covered by the cruises, and the position of ESTOC, where cruise M38/2 remained on station, is shown in Fig. 2-1.

Samples were collected by pumping seawater from a depth of approximately 5 metres through a 70 μm mesh net using the ship's fire hose system. Pumping was continuous, with samples being taken at relatively short intervals (hourly), to maintain the

viability of the foraminiferal cells. On cruise M38/2 at ESTOC a plankton net (70 μm mesh) was lowered to depth of 100 metres where it was kept for 2 minutes before being raised to the surface. Specimens obtained by this collection method were large, and remained relatively undamaged. The foraminifers at the surface tended to be small juveniles, with mature specimens being found deeper in the water column. The plankton samples were emptied from the net's cod-end and viewed under a Wild stereo microscope.

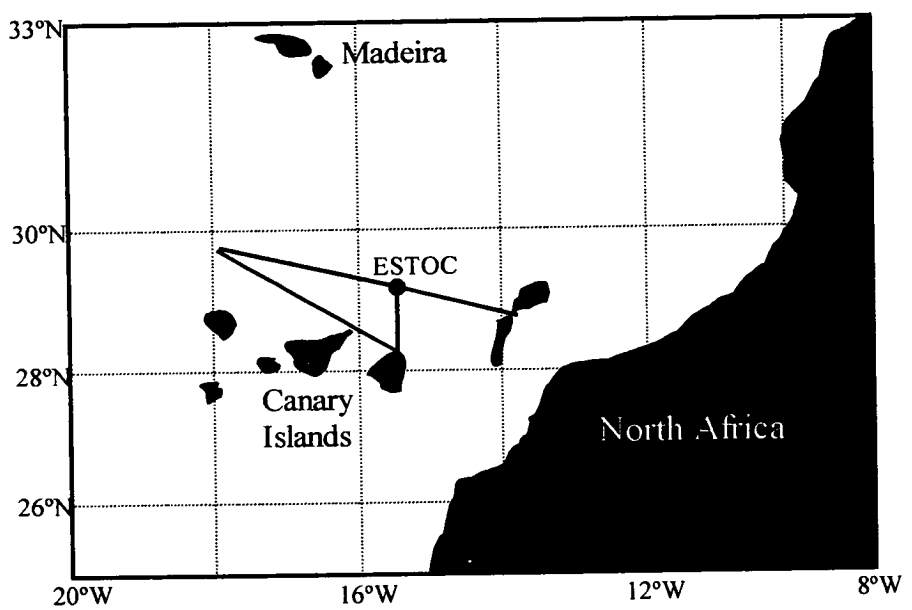


Fig. 2-1. Map indicating the location of the FS Meteor collections, north of the Canary Islands. The M37/2 transect is shown. The M38/2 station at ESTOC is represented by the circle.

2.2.2. Denmark Strait collection onboard RV PROFESSOR LOGACHEV

Sampling on this cruise was in collaboration with the Department of Earthscience at the Free University of Amsterdam. Cruise funding was provided by both the

Netherlands Foundation for Scientific Research (NWO) and the Geological Survey of Denmark and Greenland (GEUS). The cruise sampling (20th August – 3rd September 1997) commenced at a point south of Iceland (59°N/12°W) and continued along the cruise track to the south-east Greenland margin (64°N/41°W). The sampling transect is illustrated in Fig. 2-2. Planktic foraminiferal samples were collected by both pump and net methods and are outlined below.

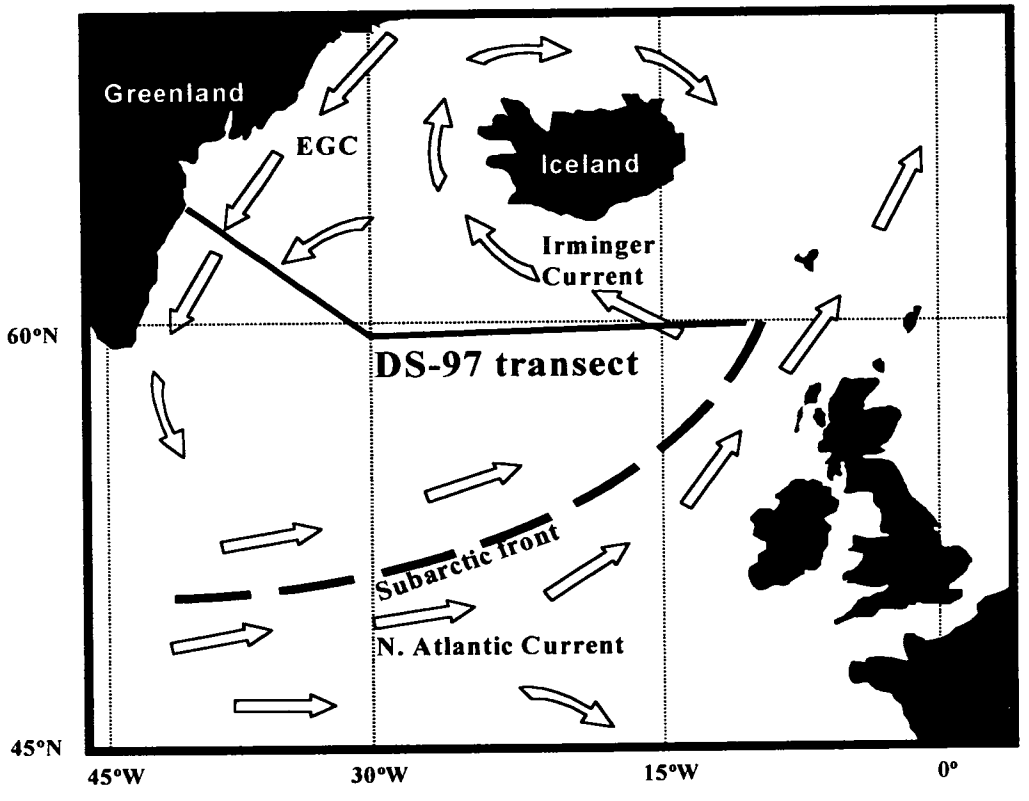


Fig. 2-2. Location map of subarctic collection onboard RV Professor Logachev. The general cruise transect and the major surface current systems are shown. The **EGC** denotes the cold East Greenland Current. The polar front is not shown due to its close proximity to the East Greenland coast.

2.2.2.1. Pump Collection

Seawater, originating from 4.5 meters subsurface, was pumped *via* the shipboard fire-hose through a 78µm mesh plankton net and cod-end, suspended in a container equipped with a flow meter. The flow was initially highly variable, but this was resolved and a steady flow of approximately 20 litres/minute was achieved. Samples were taken after periods of 1-2 hours of pumping to obtain living specimens with viable DNA. Location details of the pump collections are shown in Table 2-1. Plankton collected within the cod-end of the net was tapped off and examined under a Wild M3Z microscope.

Date	Pump start	Pump stop
20.8.97	59°57.9N/11°34.3W	60°00.7N/11°57.8W
21.8.97	59°47.5N/20°18.3W	59°47.9N/20°59.9W
22.8.97	59°38.1N/23°33.0W	59°35.5N/24°05.5W
23.8.97	59°15.6N/27°51.3W	58°56.2N/30°24.6W
24.8.97	60°13.4N/32°16.5W	60°21.7N/32°30.6W
25.8.97	62°25.0N/37°18.7W	63°01.9N/38°38.6W
26.8.97	63°47.7N/40°19.8W	63°31.5N/39°42.9W
27.8.97	62°36.3N/37°46.8W (on station all day)	
28.8.97	GPS down	63°02.4N/38°39.5W
29.8.97	63°54.1N/35°37.0W	63°56.7N/35°22.4W
30.8.97	64°19.8N/36°29.9W	64°06.9N/35°52.3W
31.8.97	63°31.4N/39°43.0W	63°28.4N/39°50.4W
01.9.97	62°25.1N/39°49.4W	62°14.6N/40°32.5W
02.9.97	62°05.6N/39°31.7W	62°01.7N/39°20.8W
03.9.97	62°22.7N/40°33.1W	62°15.8N/40°48.0W

Table 2-1. Collection details of subarctic Atlantic pump collections.

2.2.2.2. Net Collection

When the ship was on station, planktic foraminifera were obtained using a Hydro-Bios closing plankton net and cod-end with 70μm mesh. The net was lowered by winch over the side of the ship and, to ensure that the net travelled vertically down and did not interfere with the ships hull or propeller, a 50 kg weight was suspended below the cod-end. A variety of tactics were employed to obtain the best sample yield, details of which are shown in Table 2-2.

Station, Grid Ref.	Depth range (m)
3 (63°47.8N/40°19.8W)	200-0
4 (63°31.5N/39°42.7W)	250-200
4 (63°31.5N/39°43.2W	100-70
7 (63°02.4N/38°39.5W)	250-0 (twice)
8 (62°36.8N/37°46.8W)	250-0 (twice)
9 (63°56.7N/35°22.4W)	250-0
10 (64°19.8N/36°29.9W)	200-0
10 (64°19.8N/36°29.9W)	200-20
16 (62°15.8N/40°48.0W)	450-200
16 (62°15.8N/40°48.0W)	200-0
17 (62°15.9N/41°52.8W)	150-0 (twice)

Table 2-2. Details of subarctic Atlantic net collections showing the station number grid references, and the depth range through which the net was winched.

By varying the collection method, it was found that the greatest number of planktic foraminifera were collected by lowering the plankton net to the desired depth and then winching it back vertically to the surface. This method allowed the greatest volume of water to pass through the net, enabling a greater number of foraminifers to be caught. When required, the net was closed by dropping a weighted messenger down the cable to trip the net closing mechanism. This was particularly useful in areas where there

was a high concentration of phytoplankton in the upper few metres of the water column, as closure of the net a few metres below the surface prevented it from becoming clogged. The samples obtained from the net collection were of higher quality than the pump method, as all foraminifers were undamaged due to the lack of turbulence within the net. The plankton samples were emptied from the cod-end and foraminiferal specimens processed as described in section 2.2.4.

2.2.2.3. Planktic foraminiferal assemblage collection

Bulk plankton samples were obtained by pumping seawater through a plankton net as described in section 2.2.2.1. The samples were collected overnight with continuous pumping for periods of 10-12 hours. A total of 10 assemblage samples were taken. The location details of the bulk plankton assemblage collections are shown in Table 2-3.

Assemblage	Start	Stop
1	60°00.4N/1432.3W	59°49.2N/1911.6W
2	59°31.4N/2507.3W	59°15.6N/2751.3W
3	58°56.1N/3015.7W	60°13.4N/3216.5W
4	63°09.1N/3853.3W	63°43.5N/4010.0W
5	63°31.4N/3942.7W	62°36.3N/3746.9W
6	63°02.5N/3839.5W	GPS not available
7	63°56.6N/3522.4W	64°07.0N/3552.7W
8	62°43.8N/4038.0W	62°25.1N/3949.4W
9	62°07.2N/4012.4W	62°05.6N/3931.7W
10	62°23.4N/3951.1W	62°22.7N/4033.1W

Table 2-3. Location details of subarctic bulk plankton assemblage collections.

The plankton was emptied from the cod-end into a container, allowed to settle, and the excess seawater poured off. The resulting concentrate was preserved by adding

100% ethanol, with the final ethanol proportion being 60% of the total volume.

The samples were transported back to the laboratory and processed to examine the morphology of the planktic foraminifera in more detail using a scanning electron microscope (SEM) (section 2.3).

2.2.3. North Atlantic collection onboard FS POSEIDON

This cruise (P247; January 1999) was run in collaboration with the *Institute für Meereskünde* in Kiel, Germany. The initial section of this cruise was between Kiel and Ponta Delgada on the Island of San Miguel in the Azores. The ship then sailed south of the Azores, making three transects of the Azores front, before sailing to Las Palmas on Gran Canaria (Fig. 2-3).

Planktic foraminifers collected for DNA analysis and bulk plankton samples were obtained *via* the pump method as previously described (section 2.2.2.1). The collection details are shown in Table 2-4. Foraminifers were viewed under a stereo microscope and specimens were processed using the protocol described in section 2.2.4.

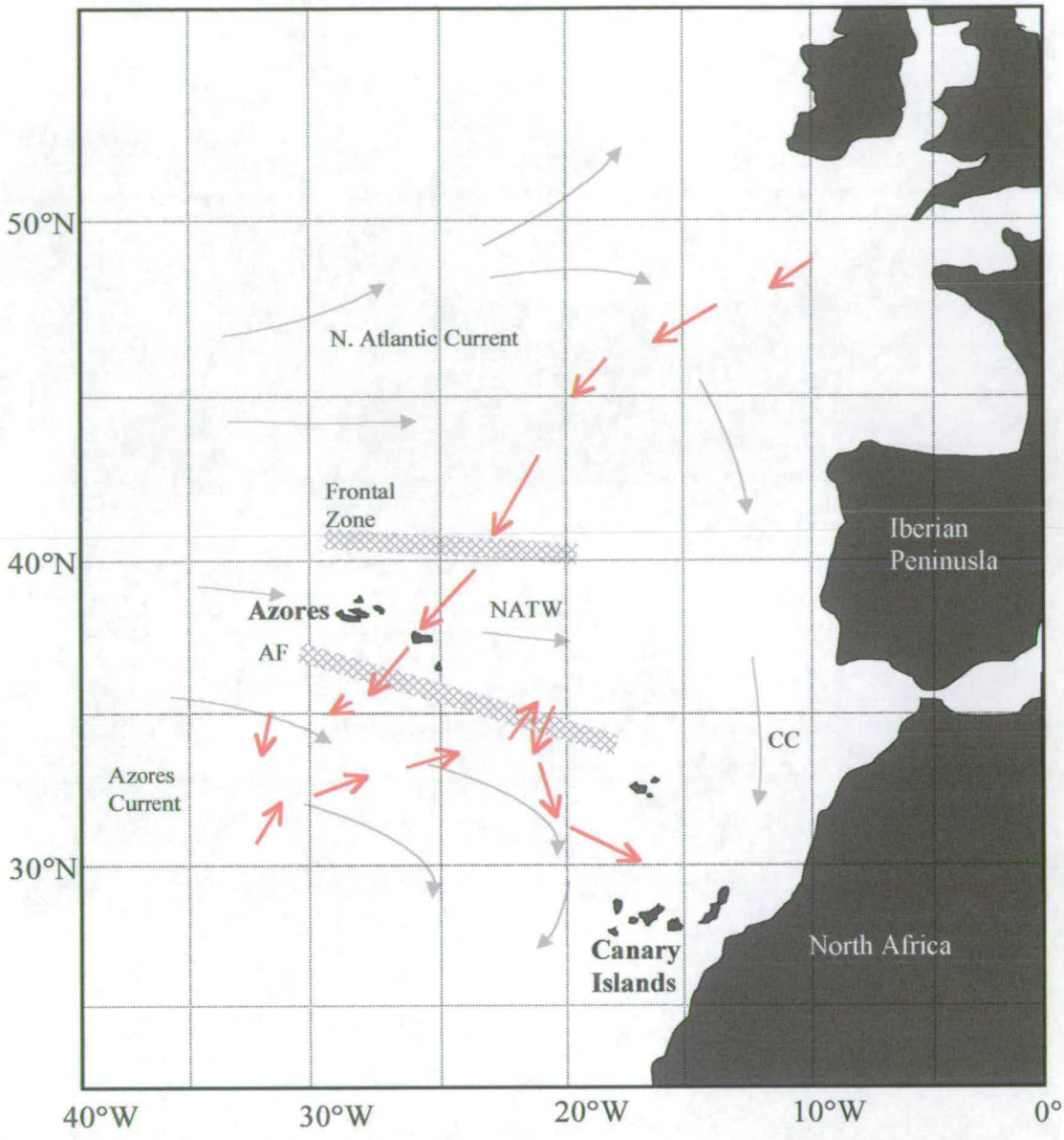


Fig. 2-3. Map of a section of the Northeast Atlantic illustrating the collection locations (red arrows) sampled during the Poseidon 247 cruise. Water mass boundaries (hatched lines) were determined by the ship's onboard thermosalinometer. The surface currents (grey arrows) are after Ottens (1991). **AF** denotes the Azores Frontal region, **CC** the Canary Current, and **NATW** the North Atlantic Transitional Water.

Date	Pump start	Temp. (°C)	Salinity (PSU)	Pump stop	Temp. (°C)	Salinity (PSU)	Hydrographic region
10.1.99	48°38.9N/10°31.5W	12.1	35.6	47°34.1N/14°27.0W	12.9	35.7	NAC
11.1.99	47°27.9N/14°49.0W	12.8	35.7	47°04.8N/16°11.4W	12.5	35.7	NAC
12.1.99	46°24.2N/18°34.3W	13.4	35.8	45°07.9N/19°50.4W	13.9	35.9	NAC
13.1.99	43°17.1N/21°26.6W	14.2	35.9	~ 42N/22W	-	-	NAC
14.1.99	39°55.3N/24°14.0W	16.7	36.1	38°43.8N/25°12.8W	16.9	36.1	NATW
16.1.99	37°28.3N/26°07.2W	17.4	36.3	36°42.4N/27°08.2W	18.4	36.5	NATW - AF
17.1.99	35°00.1N/29°11.1W	19+	36.5	35°00.1N/29°11.1W	19.3	36.5	AC
18.1.99	34°58.4N/31°03.1W	-	36.5	33°59.7N/31°31.7W	19.0	36.5	AC
19.1.99	32°06.0N/31°39.0W	20.1	-	32°29.5N/30°54.6W	19.8	36.7	AC
20.1.99	32°54.8N/29°24.2W	19.8	36.8	33°13.2N/27°55.4W	19.2	36.6	AC
21.1.99	33°36.3N/26°08.8W	19.4	36.8	33°59.9N/25°07.9W	18.8	36.6	AC
22.1.99	34°55.4N/22°48.5W	18.7	36.6	35°49.8N/20°29.7W	18.2	36.6	AF
23.1.99	35°07.7N/20°53.3W	18.6	36.7	33°57.9N/21°31.4W	19.1	36.8	AF
25.1.99	33°00.0N/22°00.0W	18.8	36.7	32°31.2N/21°10.7W	18.9	36.7	AC
26.1.99	31°27.4N/19°24.1W	18.9	36.9	30°29.9N/17°47.3W	18.9	36.8	AC

Table 2-4. Location details of each pumped sample, indicating the sea surface temperature and salinity, and the water mass zones. **NAC** denotes the North Atlantic Current water, **NATW** the North Atlantic Transitional water, and **AF** the Azores Frontal region.

2.2.4. Processing the planktic foraminifers after collection

After collection, the planktic foraminifers were picked from the plankton samples. By viewing the sample under a stereo microscope, the individual foraminifers were removed using a fine paint brush, or glass pipette, and put in a petri dish containing clean seawater. The foraminiferal specimens are often covered in debris, so they were washed in seawater to remove any adherent particles. During both the Denmark Strait collection, and the North Atlantic collection, digital images of the planktic foraminiferal specimens were recorded using a Panasonic digital video camera mounted on top of the microscope. It was hoped that this would prove useful in trying to link the morphology of each specimen to its genotype. This was unsuccessful due to the poor image quality, which

was in part due to vibration from the ship's engines, and the ships motion. Another factor that hindered this attempt was that many of the specimens collected were juvenile, so that they were too small to be "photographed" satisfactorily.

Once the foraminifer individuals had been picked and digital images taken, the DNA was protected from degradation by enzyme activity. For single cell amplification, the foraminifera were crushed in 20µl of lysis buffer, which contains 50mM of Tris (hydroxymethyl)-aminomethane, pH 8.6, 2mM of EDTA (ethylenediaminetetra-acetic acid), 0.1 % Triton X-100 and 0.5 % of sodium deoxycholate (Holzmann and Pawlowski, 1996), before incubation at 60°C for one hour and storage at below -20°C. A double strength lysis buffer (containing 100mM of Tris, pH 8.5, 4mM of EDTA, 1% of sodium-deoxycholate and 0.2 % of Triton X-100) as described by de Vargas *et al.* (1997) was tested using several subarctic samples. The higher concentration of EDTA should be better at protecting DNA from enzyme degradation, and since there was found to be no difference in the resultant PCR amplification between individuals crushed in either buffer, all samples collected during the FS Poseidon cruise were crushed in the double strength buffer.

The specimens collected during the FS Meteor cruises (M37/2 and M38/2) were not transported home frozen. Further, a number of samples (20) collected during M38/2 underwent full DNA extraction (section 2.4.2) at the Marine Institute in Telde, Gran Canaria. The low success at amplifying the DNA from these samples suggested that thawing the samples and transporting them home was a problem. In response to this, the samples collected during both the Denmark Strait and North Atlantic cruises were

transported back home in dry ice. This worked, and the amplification success was much higher than previously.

2.3. Bulk plankton sample processing and scanning electron microscopy

As foraminiferal shells (tests) are destroyed during DNA extraction, bulk plankton samples were collected to investigate possible genotype/morphotype relationships. The preserved bulk plankton samples were individually washed with water through a 63μm sediment sieve and emptied into a glass petri dish (diameter ca. 10 cm). Two different methods were employed for preparing samples for scanning electron microscopy (SEM). In the first method, the sample was dried in an oven at 65°C to complete dryness. The sample was ashed using a Fisons Instruments VG Microtech Polaron PT7300 RF Plasma Barrel Etcher. During the ashing procedure a chemical reaction occurs, which burns away the organic matter and chitin (associated with the copepods) within the sample. The calcium carbonate tests of the foraminifera are more resistant to ashing than any of the other organisms present, hence the foraminiferal tests remain intact. In the second method, the sample was also dried in an oven at 65°C. Viewing the sample under a stereo microscope, the foraminiferal tests were picked from the sample. The specimens were washed in dilute sodium hypochlorate (NaOCl), which dissolves adherent organic material. In both methods, the planktic foraminifera were finally washed using distilled water to remove adherent debris.

The individual planktic foraminifera to be viewed under the SEM were mounted on a carbon-backed stub. Using a stereo microscope, the foraminifers were carefully

aligned in rows on the stub, with each specimen orientated according to the side to be photographed (umbilical or trochospiral). The stub was sputter-coated with a gold coat (~ 80 nm thick) using a Bio-Rad Polaron Division SEM Coating System. The carbon-backed stub and the gold coating are required to ensure that the samples are conductive to an electrical current within the electron microscope. Digital images of the individual foraminifera were recorded using a Philips XL 30 CP scanning electron microscope. Each image was saved as a *.tif file, for easier manipulation within PC-based drawing packages.

2.4. Molecular biology methodology

The first attempts to extract and amplify foraminiferal rDNA were made by Langer *et al.* (1993) and Wray *et al.* (1993). The main problem they encountered was the difficulty in isolating foraminiferal rDNA from contaminant rDNA. The contaminants can be DNA from prey which the foraminifer has ingested, or DNA from algal symbionts which many of the planktic foraminifers possess. Wray *et al.* (1993) attempted to overcome the problem by using non-symbiont bearing benthic foraminifers but, as was later shown by Darling *et al.* (1996b) and Pawlowski *et al.* (1996), the amplified products were still contaminants, being too small to be of foraminiferal origin. In reviewing techniques for the isolation and amplification of genomic DNA from planktic foraminifera, Darling *et al.* (1996a) showed that fingerprinting techniques were not viable due to the small quantities of DNA involved, and that a PCR-based DNA fingerprinting technique known as RAPD (Randomly Amplified Polymorphic DNA) was unsuccessful.

It was suggested that instead of continuing fingerprinting approaches, amplification of the highly conserved small subunit (SSU) rRNA gene, would prove a more successful approach. To aid the identification of foraminiferal DNA fragments, Darling *et al.* (1996a,b) amplified rDNA from specimens that had reached their reproductive stage (gametogenesis). At this point in a planktic foraminifer's life cycle, specimens are observed to consume their symbionts to gain sufficient energy for DNA replication and are also observed to dispose of any remaining intracellular particles by cytoplasmic streaming. Gametogenic specimens may contain up to 250,000 copies of the genomic DNA (Spindler *et al.*, 1978; Bé *et al.*, 1983), and therefore a reduced proportion of contaminant DNA, increasing the probability that foraminiferal DNA will be amplified. Once the correct foraminiferal fragments had been identified (see Darling *et al.*, 1996b, Wade *et al.*, 1996), the isolation and amplification of foraminiferal DNA improved since specific primers were designed (section 2.4.2.2). In more recent years, it has been possible to amplify foraminiferal DNA from non-gametogenic specimens, due to the improvement of reagents, such as Taq DNA polymerase, used in the amplification of DNA.

Currently, there are two approaches used in the isolation and amplification of foraminiferal rDNA. The first is by PCR amplification and direct Taq cycle sequencing (Darling *et al.*, 1996b, 1997). The second is by PCR amplification followed by the ligation and cloning of the amplified template, and sequencing of the clone (e.g. de Vargas *et al.*, 1997; Pawlowski *et al.*, 1997). The disadvantages of the first method are that the sample must be homozygous for the sequence of interest, the sequence of the 5'

and 3' terminal region of the fragment must be known, and the cost of Taq DNA polymerase is high. The first method has the advantage of being faster than the second method, and direct sequencing of PCR products reduces the problems associated with Taq-induced amplification errors, since they will appear as ambiguities in the sequence. Further, the first method will show whether sequence length polymorphism exists within a single specimen, since a length difference of only 1 nucleotide will result in multiple templates within the nucleotide sequence. To identify sequence length polymorphism using the second method, many clones would need to be sequenced. Fortunately, intraspecific sequence length polymorphism has not been a problem in the amplification of planktic foraminiferal SSU rDNA. For a comparison of the advantages/disadvantages of these two techniques see Hillis *et al.* (1990). Following the protocols employed within our laboratory, I used the PCR amplification and direct sequencing method for investigating planktic foraminiferal SSU rDNA.

2.4.1. DNA Extraction

It was found that direct PCR amplification from lysis buffer was unsuccessful due to the inhibiting effects of the buffer constituents (see section 2.4.2.4). By extracting the DNA prior to amplification, these inhibiting reagents and debris from the lysis crush are removed, and the nucleases from the foraminiferal cell, which would otherwise degrade the DNA, are inactivated. The full DNA extraction is described in the following protocol (after Darling *et al.*, 1996b).

1. Individual lysis crushes are thawed and the volume brought up to 250 µl with 0.1 M

EDTA + 0.25 % SDS (lauryl sulphate) w/v.

2. Samples are incubated in the presence of proteinase K (0.125µg / 5µl) at 65°C for 1-3 hours.

3. 75 µl of 3.5 M sodium chloride (NaCl) is then added together with 33 µl of CTAB (cetyltrimethylammonium bromide) (100 ml contains 10g CTAB, 20ml of 3.5M NaCl and 80ml of distilled H_20) and 4 µl of tRNA (1µg/4µl).

4. The mixture is incubated at 65°C for 1 hour.

5. An equal volume of chloroform is then added and mixed gently for 5 minutes.

6. Centrifuge at 10,000g for 6 minutes.

7. The supernatant is removed without disturbing the interface and an equal volume of phenol : chloroform : isoamyl alcohol (25:24:1) is added. Mix gently for 5 minutes and centrifuge at 10,000g for 6 minutes.

8. The supernatant is removed without disturbing the interface and an equal volume of chloroform : isoamyl alcohol (24:1) is added. The tubes are again mixed gently for 5 minutes and centrifuged for 6 minutes.

9. The supernatant is removed and 2 volumes of 100 % ethanol are added. The tubes are placed at -20°C for over 1 hour (usually overnight).

10. The tubes are centrifuged at 10,000g for 20 minutes, the position of the DNA pellet is identified, and the alcohol is carefully aspirated.

11. The pellet is washed in 80 % ethanol, centrifuged for 10 minutes, and the alcohol aspirated. The pellet is then dried at 60°C on a hot-block.

12. The pellet is resuspended in 50 µl of sterile water and stored at –20°C.

The EDTA provides the DNA with protection from enzyme activity and the SDS acts as a surfactant removing oils from the samples. Proteinase K degrades the nucleases that would breakdown the DNA. Transfer RNA (tRNA) acts as a carrier to aid the precipitation of small volumes of DNA. This is necessary since the amount of genomic DNA present in one cell is very small. Phenol destroys any remaining proteins. The chloroform removes unwanted material by its solvent action, and also any residual phenol. Isoamyl alcohol is used in the process since it provides a discrete interface between the supernatant and the unwanted material below, aiding the removal of the supernatant. Alcohol is present to precipitate the DNA from solution and to remove excess salt residue from the side of the microcentrifuge tube. This is necessary since excess salt carried through into the PCR may upset the balance of the reaction, thus inhibiting DNA amplification.

2.4.2. Amplification of ribosomal DNA using Polymerase Chain Reaction (PCR)

The PCR has revolutionised modern molecular biology. It is the process by which a DNA fragment of unknown sequence can be amplified to create millions of identical copies of the fragment (Mullis and Faloona, 1987; Saiki *et al.*, 1988), allowing the target DNA to be detected and, in this case, sequenced. The reaction involves the DNA template being added to a chemical mixture containing short strands of DNA to prime the reaction, a DNA polymerising enzyme, and the nucleotides required to replicate the template. The reaction mix is repeatedly heated and cooled in a series of steps, which

allows the different stages of the amplification to take place. The PCR mixture contains the following:

1. DNA template. The primary PCR uses genomic DNA, and subsequent reactions use the DNA fragment amplified in the primary reaction. Amplification can only take place if the ends of the sequence are known (see below).

2. Primers (10μM). Primers are short fragments of DNA (usually 20 or more bases) designed to complement the known sequence at each end of the unknown fragment to be amplified. In each PCR mix, two primers, one at each end of the DNA strand, are used. They anneal to the template DNA at their corresponding complementary position and form the starting points from which the DNA polymerisation takes place.

3. Taq (*Thermus aquaticus*) DNA polymerase (5 units per 100μl reaction). This enzyme is the basic driving force behind the PCR as it replicates the complementary strands of the target sequence starting from the two primer sites (Hillis *et al.*, 1990). It was developed from the DNA polymerase found in thermophilic bacteria that live in hot springs. Its thermostable nature allows it to tolerate the repeated heating to the high temperatures required to denature the DNA, which is a feature of PCR technology.

4. Deoxyribonucleoside triphosphates (dNTPs, 2.2 mM). The four nucleotides, adenine, guanine, thiamine and cytosine, are required to build the copies of the DNA template.

5. PCR buffer containing magnesium chloride ($MgCl_2$, 3.5 mM). The Mg^{2+} ions within the buffer make complexes with the dNTPs, primers and DNA templates. Too low Mg^{2+} ion concentration will result in low PCR yields.

6. Sterile water (diethylpyrocarbonate (depc) treated). The depc treatment destroys nucleases that may exist within the distilled water. After sterilisation, the treated water must be autoclaved to destroy the depc since it can cause failure of subsequent PCRs.

2.4.2.1. The PCR cycle

There are three stages to the PCR process, each of which have a specific temperature and running time. The GeneE PCR machine (Techne), used for this project, cycles through these stages automatically following a set program.

The three stages are as follows:

1. Denaturing stage - this has a high temperature (94°C) to separate the DNA into single strands.

2. Annealing stage – this is the stage in which the primers attach to the target DNA template at the corresponding complementary position. Annealing temperatures vary with different primers and section 2.4.2.2. below outlines these features, together with primer design and function.

3. Extension stage - the DNA strands are produced by extension from the last nucleotide of the primer. The nucleotides (dNTPs) in the reaction mix are added in the correct order to the complementary strand of DNA.

Each of the stages in the PCR are illustrated together in Fig. 2-4, and Table 2-5 shows the temperatures and duration of the individual PCR stages.

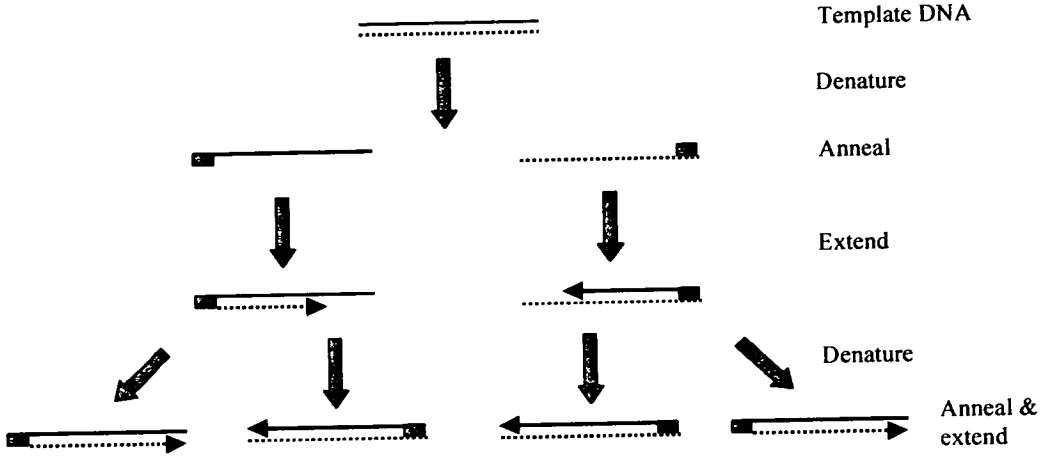


Fig. 2-4. Pictorial representation of the PCR process. The diagram starts with the template DNA obtained from the DNA extraction. This is denatured, and the primer (▬) anneals to the single strand of foraminiferal SSU rDNA before extending along its length, thus creating an identical copy. This process is repeated to create an exponential number of copies of the target region of the SSU rDNA required.

Cycles	Cycle (x1)	Cycle (x20 to x35)	Cycle (x1)
Denaturing	2 min at 94°C	25 sec at 94°C	7 min at 68°C
Annealing	5 min at 48 – 55°C	35 sec at 48°C	-
Extension	4 min at 72°C	2 min at 72°C	-

Table 2-5. The specific temperatures and duration for each stage of the PCR process used in this study.

2.4.2.2. Primer design and function

The following terms are used to explain the orientation of a DNA molecule: 5' (left region of molecule), 3' (right region of molecule), sense (forward strand of DNA) and antisense (reverse strand of DNA). The terms, sense and antisense, are also used to

describe the direction in which a primer amplifies (Table 2-6). In each PCR two primers are present, corresponding to positions at the 3' and 5' end of the template DNA fragment to be amplified. Primers are usually described as being specific or universal (non-specific). The specific primers tend to be longer and are closely matched to the target template sequence. These primers will preferentially amplify the foraminiferal DNA rather than any other DNA that may be in the reaction. The universal primers are typically short in length and will amplify any DNA in the reaction medium.

Each primer has a temperature tolerance limit for annealing, known as the thermal maximum (Tm), above which the primer does not anneal. The thermal maximum of a primer is a function of its nucleotide sequence, and is calculated as follows:

$$Tm\ (°C) = 4 \times (\text{number of G and C}) + 2 \times (\text{number of A and T})$$

e.g. primer N5 has a nucleotide sequence of AAC TTA AAG GAA TTG ACG GAA G

Therefore, $Tm = (4 \times 8) + (14 \times 2) = \mathbf{60°C}$

The specificity of the primer annealing can be altered by varying the annealing temperature. The higher the annealing temperature (towards its thermal maximum), the higher the specificity of the primer. Some samples may be difficult to amplify due to a mismatch between the primer sequence and the annealing site on the target DNA. This can be overcome by decreasing the annealing temperature so that the PCR is less specific, hence promoting more amplification within the reaction. Low annealing temperatures are avoided where possible since it can cause multiple priming, i.e. where more than one fragment is amplified within a single reaction.

To amplify the ~1000 bp 3' terminal region of the SSU rRNA gene within planktic foraminifera the following primers (Table 2-6) have been used, after White *et al.* (1990) and Darling *et al.* (1996b). The positions of these primer sites on the 3' terminal region of the SSU rRNA gene are shown in Fig. 2-5.

Primer	Direction	Sequence
C5	Sense	GTA GTA TGC ACG AAG TGT GA
N5	Sense	AAC TTA AAG GAA TTG ACG GAA G
NN6	Antisense	GCA TCA CAG ACC TGT TAT TGC C
FS3	Sense	GTG ATC TGT CTG CTT AAT TGC
FS4	Sense	GAA CGC AAC GGA CGT G
N7	Sense	GGC AAT AAC AGG TCT GTG ATG C
138	Antisense	TGA TCC TTC TGC AGG TTC ACC TAC
N8	Antisense	TCC GCA GGT TCA CCT ACG GA

Table 2-6. Details of the SSU rDNA primers used for PCR. The amplification direction and the specific nucleotide sequence for each primer is shown. After White *et al.* (1990) and Darling *et al.* (1996b).

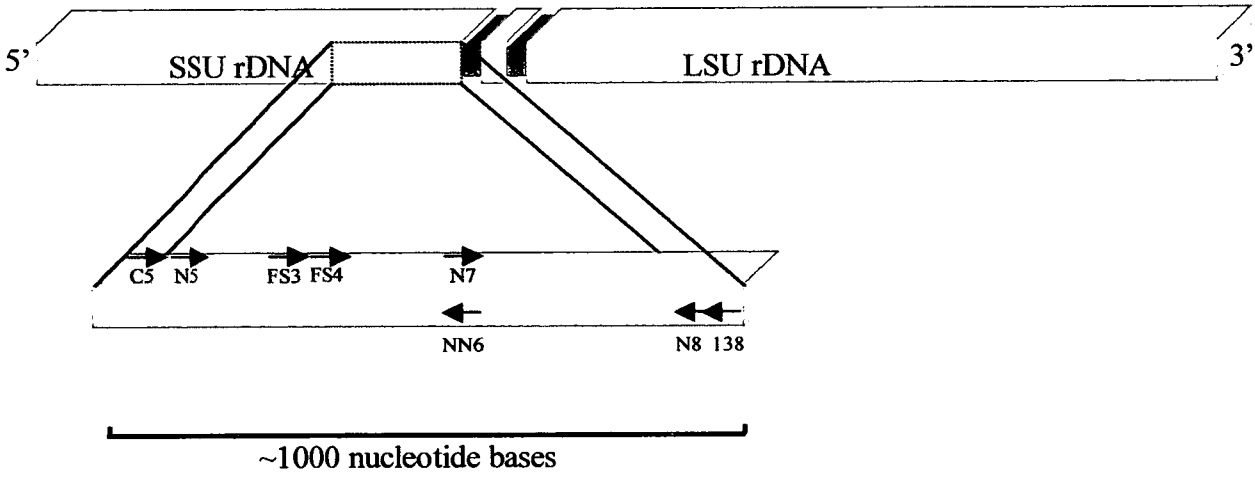


Fig. 2-5. Diagram of rRNA array. For an explanation of the array refer to Chapter 1. The expanded segment illustrates the ~1000 bp 3' terminal region of the SSU rRNA gene amplified. The positions of the primer sites are shown.

The procedure starts with a primary PCR amplification from the extracted

genomic DNA, using primer pairs C5/138 which amplify the section of the gene under study (~1000 bp). The primary PCR primers are universal and will amplify any DNA within the reaction, including symbiotic algal DNA.

Subsequent secondary PCRs reduce the size of the target fragment to a size suitable for sequencing by "nesting" the reaction with the inner primers N5 / NN6 (~500 bp), N7 / 138 (~500 bp) or FS3 / 138 (~700 bp). A contiguous sequence can then be assembled from the separate fragments.

2.4.2.3. Taq enzyme

A number of thermostable enzymes are available for PCRs. Taq DNA polymerase from the following manufacturers have been used: Promega, Helena Biosciences, Perkin Elmer and Sigma. It was found that the Helena Biosciences Taq Supreme gave the greatest amplification yields. This may be due to the enzyme having a relatively high activity. The activity is measured as the concentration of dNTPs incorporated into the reaction per unit of time. For example, the Helena Biosciences Taq Supreme defines 1 unit incorporating 60nmoles/30 minute [Helena Biosciences data sheet, EP0302], where as Promega Taq defines 1 unit as incorporating 10nmoles/30 minute [Promega data sheet, M2861].

2.4.2.4. Inhibition of the PCR by lysis buffer

The foraminiferal DNA can be directly amplified from the lysis crush, as described by Pawlowski *et al.* (1994) and Holzmann and Pawlowski (1996), but it has been found

that the success rate was very low. It is known that the EDTA component of lysis buffer is an inhibitor of the PCR. EDTA inhibition was tested by titration with a set of reactions using DNA from a Caribbean *G. siphonifera* Type I specimen. The experiment was set up so that in the primary PCR, half of the reactions contained lysis buffer at ~25% of the final PCR volume. The second round of amplification was done as a standard reaction and the PCR products were separated by gel electrophoresis. As Fig. 2-6 shows, the lysis buffer greatly reduced the amplification yield of the PCRs. This is probably due to the magnesium chloride within the PCR buffer cholating with the EDTA in the lysis buffer, thus reducing the available $MgCl_2$ for making complexes with the dNTPs, primers and DNA templates, and therefore inhibiting the reactions. To counteract the inhibition effects of the lysis buffer within the PCR, the DNA from the crushed foraminiferal cell was extracted fully according to the protocol described above (section 2.4.1.).

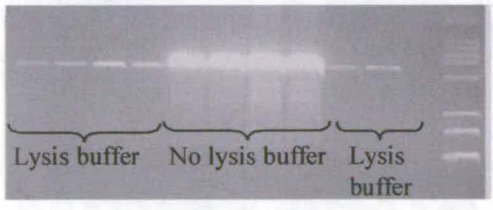


Fig. 2-6. Gel electrophoresis indicating the inhibition effect of lysis buffer. The brighter bands indicate high amplification yields, and the weak bands indicate inhibited amplification. The SSU rDNA region was amplified using a primer combination of N5 and NN6 (Fig. 2-5).

2.4.2.5. Effect on PCR by varying magnesium chloride concentration

As the lysis buffer experiment showed that the available concentration of $MgCl_2$ within the PCR has a profound affect on the success of the reaction, an experiment was set up to test the effect of varying $MgCl_2$ within the PCRs. A series of reactions were set up containing DNA, known to have amplified successfully, and $MgCl_2$ in concentrations ranging from 2.5-4.0 mM. Each reaction was duplicated for reproducibility. The results are shown in Fig. 2-7. Only the low $MgCl_2$ concentration of 1.5 mM failed to amplify any DNA. For the other reactions there was no significant difference between the resultant amplification so the Helena Biosciences PCR buffer, which gives a $MgCl_2$ concentration of 3.5 mM, was continued to be used.

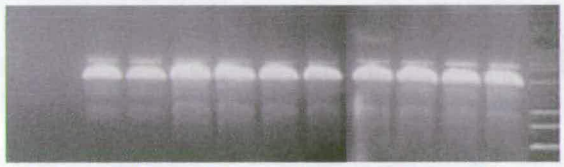


Fig. 2-7. Comparison of $MgCl_2$ concentrations within the PCR reaction mix. The concentration ranges from 1.5-4.0 mM. The larger bands are planktic foraminifera SSU rDNA, with the smaller bands being algal rDNA. The SSU rDNA region was amplified using a primer combination of FS4 and N8 (Fig. 2-5).

2.4.2.6. Effect on PCR by varying conditions

The PCR conditions used are those described by Darling *et al.* (1996). Slight modifications are employed depending on the nature of success of individual samples. For example, the number of cycles can be varied according to how well an individual sample is amplifying, with less cycles being required when a particular sample amplifies easily.

Further, the annealing temperature can be varied between 48°C and 52°C, with the higher temperature giving the primers more specific annealing to the template DNA. If a sample is proving difficult to amplify then a lower annealing temperature (and/or more PCR cycles) often helps amplification. Much of the PCR condition variations employed are on a matter of trial and error to find what works best with the individual specimens.

2.4.3. Separation of amplified DNA fragments by agarose gel electrophoresis

The PCR products are separated by agarose gel electrophoresis. The PCR products are loaded into wells formed within a 1 % agarose gel stained with ethidium bromide (0.5 μg/ml) to illuminate the DNA. The gel is made by dissolving agarose (final concentration of 1 % w/v) in TBE (Tris-borate-EDTA) buffer (5 litres contains 540g of TRIS, 275g of boric acid, and 200ml of 0.05M EDTA at pH 8.0). The gel sits within a tank of 1 % TBE buffer (containing 0.5 μg/ml ethidium bromide). A current (60-100 volts) is passed through the agarose gel using electrodes located at either end of the tank. The current causes the DNA to migrate through the gel, with the smallest fragments migrating furthest, thus resulting in separation of the different sized fragments. The ethidium bromide causes the DNA to fluoresce under ultraviolet light and the DNA fragments can then be observed as discrete bands on the gel (Fig. 2-8). In order to view the PCR products, 8 μl of the PCR product is run out on a low grade agarose gel. The size of the DNA fragment is determined by running a calibration marker (pGem, Promega) alongside the samples (Fig. 2-8). It contains a ladder of 15 fragments of known size which allows direct comparison and identification of the target band. To recover the

PCR products for clean-up and sequencing, they are run on a SeaPlaque low melting point (LMP) gel in volumes of between 40-80 µl. The desired bands of DNA are excised with a clean scalpel blade and placed in a 1.5 ml micro-centrifuge tube.

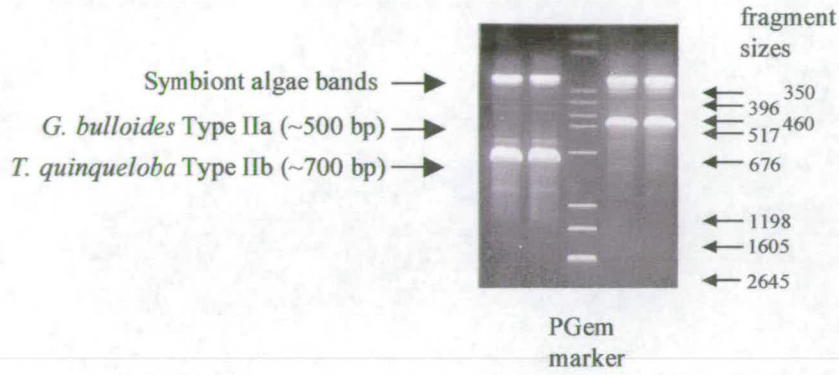


Fig. 2-8. Agarose gel electrophoresis. Two planktic foraminiferal bands are shown. Each was amplified using the primer combination of N5 and NN6. The PGem marker ladder with different fragment sizes is also shown.

2.4.4. DNA clean-up for sequencing

The microcentrifuge tubes containing the excised fragments are placed on a heating block at 70°C for approximately 20 minutes, or until the gel has melted. The samples are then cleaned-up to remove the DNA from the agarose using a Promega Wizard PCR Prep DNA purification kit according to the manufacturers protocol.

1. 1 ml of guanidine thiocyanate resin is added to the melted gel and mixed for 20 seconds.

2. The gel and resin mix is syringed through a minicolumn and 2 ml of 80% isopropanol is eluted through the minicolumn, to remove any remaining agarose/resin mix.

3. The DNA binds to the inside of the minicolumn, which is spun dry for 2 minutes at 10,000 g.

4. 40 µl of warm sterile water is then added to the minicolumn and left for 1 minute. The DNA is then eluted from the minicolumn by pulse spinning for 20 seconds, leaving the DNA in a concentrated solution.

5. The DNA is then stored at -20°C ready for Taq cycle sequencing.

2.4.5. Taq Cycle Sequencing

The culmination of the previous steps is to sequence the cleaned-up fragment of DNA. The sequencing procedure has a number of components. Different length single strands of DNA are amplified and each terminal nucleotide is labelled with one of four dyes, corresponding to the specific terminal nucleotide type (A, T, G or C). The products are precipitated in ethanol and pelleted by centrifuging. The dye labelled DNA is run through an acrylamide gel, clamped vertically in the automatic Applied Biosystems (ABI) 373A DNA sequencing machine. A current (2500v) is passed through the gel causing the amplification products aligned along the top of the gel to migrate downwards through the gel. Towards the base of the gel a laser scans the migrating labelled DNA fragments causing each of the dyes to fluoresce. Each of the 4 dye-labelled nucleotides have a specific emissivity (Table 2-7), and it is this signal which is detected by the sequencing machine. The analysis software translates the emissivity signal resulting in a series of coloured peaks which correspond to the DNA sequence.

Nucleotide	Colour	Wavelength (nm)
G	Blue	531
A	Green	560
T	Yellow	580
C	Red	610

Table 2-7. Details of each nucleotide's corresponding dye terminator colour and emissivity wavelength.

2.4.5.1. Taq cycle sequencing reactions

This involves amplifying the DNA fragments using the Gene E PCR machine. The 10 µl reactions differ from the PCR reactions previously performed, as only one primer is added in each reaction. The primers used for sequencing are the same ones employed during normal PCRs (Fig. 2-5). The primer anneals to its complementary strand and extends from there resulting in single strands being amplified in each cycle. During the PCR the growing nucleotide chain is simultaneously terminated and labelled with the specific dye corresponding to the terminal nucleotide. Terminations occur at random so at the end of the PCR the reaction contains an array of different fragment lengths with labelled terminal nucleotides which, collectively, correspond to each base of the amplified DNA strand.

Each 10 µl reaction contains:

- 4 µl of buffer (contains 200mM Tris, pH 9.0 and 5mM $MgCl_2$). The buffer is needed to balance the reaction, since only one quarter of Perkin Elmers recommended quantity of reaction mix is used.

- 2 µl of primer (final concentration of 3.2 pM).

- 2 µl of DNA template.

- 2 µl of Perkin Elmer ready reaction mix (contains dRhodamine dye terminators, dNTPs and Taq enzyme). A titration determined that 2 µl was a sufficient quantity (section 2.4.5.4), but the manufacturer's recommended quantity is 8 µl.

The reaction mix is overlayed with a drop of mineral oil to prevent evaporation during the PCR. The reaction mix and tubes must be covered by tin foil throughout the procedure as the reaction mix is photo-sensitive. The sequencing cycle conditions are as follows:

Denature:	95°C for 30 seconds	25 cycles
Anneal:	50°C for 15 seconds	
Extend:	60°C for 4 minutes	

2.4.5.2. Ethanol precipitation of DNA

Following the cycle sequencing reactions, the DNA is precipitated out of solution by transferring the 10µl reaction to a microcentrifuge tube containing 50 µl of 95 % ethanol and 3.0 µl of 2M sodium acetate, pH 4.6. The contents of the microcentrifuge tubes are vortexed and placed on ice for 10 minutes before being centrifuged at 14,000g for 20 minutes. A DNA pellet can often be visualised at this stage. The alcohol solution is aspirated and the DNA pellet rinsed with a further 200 µl of 70 % ethanol to remove any residual salt. The ethanol is aspirated and the DNA pellet is dried for 2 minutes at 90°C.

The dry DNA pellets may be stored at –20°C for up to 2 weeks. Prior to loading, 2 µl of formamide EDTA (FE) is added to each of the samples, which are subsequently vortexed then pulse spun. The FE has a ratio of 5:1 de-ionised formamide to EDTA (50mM). To ensure that the labelled DNA within each sample is single stranded, they are heated at 90°C for 2 minutes and snap-chilled on ice before being loaded into wells on the acrylamide gel within the sequencing machine.

2.4.5.3. Acrylamide gel

The gel is made up to a volume of 80 ml as follows (after manufacturer's protocol):

- 40 g of urea.
- 9.5 ml of 40 % acrylamide.
- 27 ml of distilled water.
- 1 g of mix bed resin (amberlite).

1. Two electrophoresis glass plates, specially designed for the sequencer, are cleaned and clamped together with clips in a horizontal position on top of two boxes to elevate above the work bench. Two side spacers ensure there is a slight gap between the plates.

2. The gel mixture is dissolved and filtered with 8ml of filtered 10x TBE buffer (1 litre contains 107.8g of TRIS, 55.0g of boric acid, and 8.2g of EDTA). The mixture is also degassed to eliminate air bubbles in the gel.

3. 400 µl of ammonium persulfate (APS) (0.1g of APS in 0.95ml sterile water)

and 45 µl of TEMED (N, N, N', N'- Tetramethylethylenediamine) is then added.

4. The gel mix is injected between the plates using a 50 ml syringe, and a spacer inserted between the plates to create a gap at the top end of the gel to provide space for loading the samples. The gel takes about 1-2 hours to set, after which time it can then be raised from the horizontal and the top spacer is removed. The space at the top of the gel is washed in distilled water and the outside of the plates washed in hot water to remove excess acrylamide gel. The gel plates are then allowed to air dry. Once the gel has been fitted into the 373A DNA sequencing machine, a comb is introduced into the top to create wells into which the samples are loaded. This can be either a 48 or 64 well comb depending on the number of samples to run.

2.4.5.4. Taq cycle sequencing optimisation

The dye terminators used until July 1998 were discontinued by Applied Biosystems, and replaced by a new reaction mixture which uses dRhodamine dyes. The dRhodamine reaction mix has a slightly different chemistry and when tried initially the nucleotide peak heights at the start of the sequence (the first 50 bases) were very low and could not be read by the sequence analysis software. The manufacturers recommendations are to use 8µl of the ready reaction mix. However, when using the previous kit, we found that using only 6µl of ready reaction mix and making up the difference with sterile water was actually advantageous. Using the reduced amount with the new kits however, may have upset the magnesium balance within the new reaction mix, affecting the amplification and resulting in low peak heights. To counteract this

problem, Applied Biosystems suggested using a buffer consisting of 200mM Tris, pH 9.0 and 5mM magnesium chloride to make up the difference in the reaction mix. They also suggested that the problem may be due to excess DNA. The amount of DNA should be between 5 and 20 ng, and up until that point the quantity of DNA added to the sequencing reaction was being estimated by running the DNA to be sequenced on an agarose gel and estimating by eye. For these reasons it was decided to estimate the sufficient amount of DNA added to the reaction mix accurately, to determine whether it was contributing to the low peak heights. The ratio of reaction mix/buffer was also varied.

A sequence run was set up using two specimens from the subarctic Atlantic. Both of these are right coiling *Neogloboquadrina pachyderma* and had previously sequenced successfully, apart from the low peak heights at the start. The primer used within the sequence reaction was NN6. The reagent combinations are shown in Table 2-8.

Lane	Reaction mix, µl	Primer, µl	DNA, µl	dH_20, µl	Buffer, µl
1	8	4	5	3	0
2	8	4	5	0	3
3	6	4	5	5	0
4	6	4	5	0	5
5	6	4	5	5	0
6	6	4	5	0	5
7	6	4	4	6	0
8	6	4	4	0	6
9	6	4	3	7	0
10	6	4	3	0	7
11	4	4	5	7	0
12	4	4	5	0	7

Table 2-8. Reaction combinations testing reagent quantities, in an attempt to solve the low peak heights at the beginning of the sequences. Each of the reactions had a final volume of 20 µl.

The fragments successfully sequenced with clear peaks and low background. However, none of the combinations improved starting peak heights. It was however notable, that the ready reaction mix volume could be reduced by at least half without any difference being observed in the quality of the DNA sequences. Also, the reduction of DNA template added to the reaction mix did not reduce the quality or quantity of amplified product produced.

To investigate how much the concentration of DNA template and dRhodamine reaction mix within sequencing reactions could be reduced, a series of sequencing reactions was again performed with varying quantities of reagents and DNA (Table 2-9).

Lane	Reaction mix, µl	DNA, µl	Primer, µl	Buffer, µl
1	4	6	4	6
2	4	4	4	8
3	3	6	4	7
4	3	4	4	9
5	2	6	4	8
6	2	4	4	10
7	1	6	4	9
8	1	4	4	11
9	0.5	6	4	9.5
10	0.5	4	4	11.5
11	2	3	2	3
12	2	2	2	4
13	1.5	3	2	3.5
14	1.5	2	2	4.5

Table 2-9. Reaction combinations testing reagent quantities. Lanes 1-10 had final volumes of 20 µl, and lanes 11-14 had final volumes of 10 µl.

Unexpectedly, the best sequences were obtained by using 10µl reactions. The peak heights were readable from the beginning of the sequence. It was found that by using a quarter of the reaction mix in a 10µl cycle sequencing reaction excellent results are obtained. This fortunately has important consequences for cost, since 4 times as many

sequencing reactions can be obtained from a sequencing kit.

2.4.5.5. Automated sequence collection

Throughout the sequencing run (12 hours), data is collected and processed to provide an image of the gel (Fig. 2-9) with the four dye colours depicting the nucleotide sequence displayed for each sample. Lines passing down through each sample lane are also displayed. These are computer generated tracks which the software has calculated to be the line of best signal through each sample lane and from which the sequence data is generated.

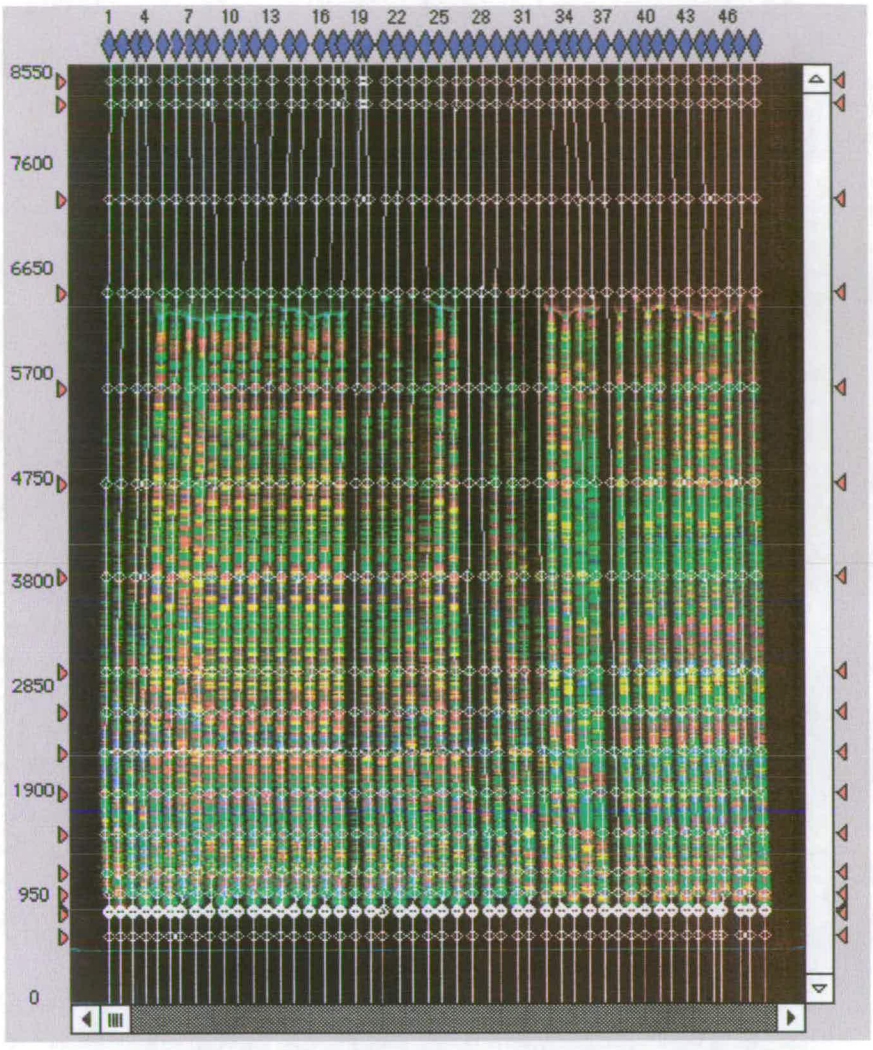


Fig. 2-9. Illustration of sequencing gel with 48 sample lanes.

The nucleotide sequence for each lane calculated by the analysis software, is known as an electropherogram and is displayed as a series of peaks, each corresponding to a specific nucleotide base. An illustration is shown in Fig. 2-10.

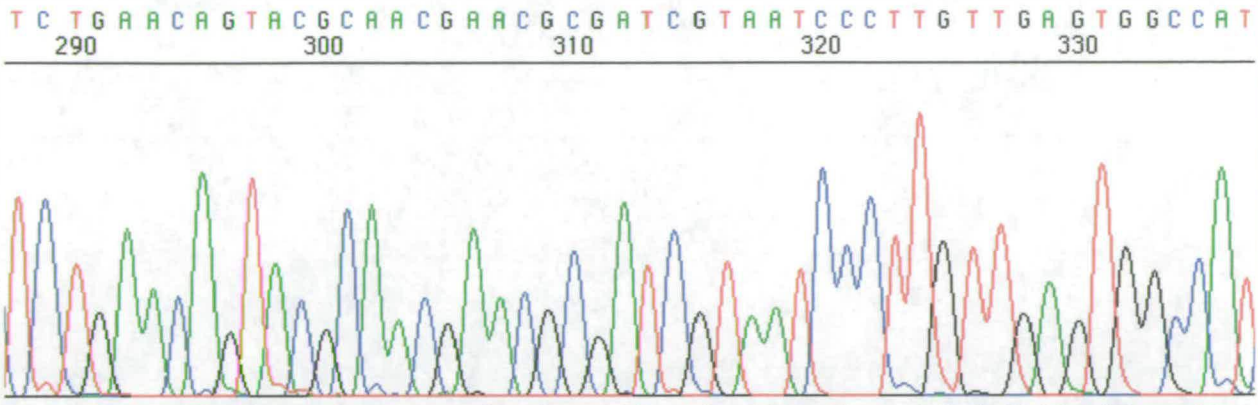


Fig. 2-10. Electropherogram showing a 50 nucleotide region of *G. bulloides* (Type Ib) from the North Atlantic.

2.5. Nucleotide sequence editing

Once sufficient fragment duplicates have been generated they are transferred to the SUN workstation for processing. Each transferred file contains the sense and antisense sequence fragments of one individual which overlap to give the full ~1000 base pair sequence required. The sequence data files undergo preliminary editing within TED using SEQPROCESS. During automated DNA sequencing, the strength of the emissivity signal tends to tail off towards the end of the sequence. This results in unread nucleotide positions, or gaps within the sequence. This package allows, where necessary, the ends of the sequences to be cut off to minimise the amount of poor data on each sequence file. The processed file is saved as a *.ted file and the sequence data can then be edited using the program SEQEDIT.

In this program the sense and antisense fragments of an individual foraminiferal specimen are brought together within a window. After sequencing the antisense

fragments (N6, 138) are actually the wrong way round and back to front (Fig. 2-11), but the software corrects this by reversing and complementing the antisense sequences (Fig. 2-11).

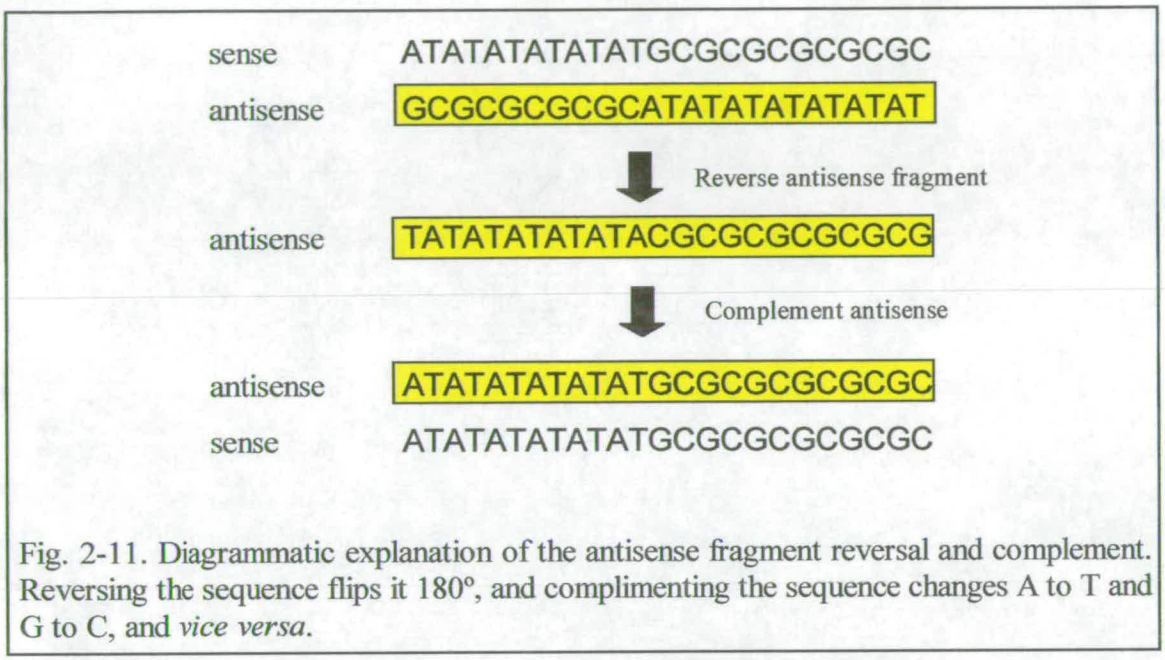


Fig. 2-11. Diagrammatic explanation of the antisense fragment reversal and complement. Reversing the sequence flips it 180°, and complimenting the sequence changes A to T and G to C, and *vice versa*.

There are sometimes a few errors or gaps within the sequences due to nucleotides having not been read by the collection software of the ABI 373A Automatic DNA sequencer. However, this is usually overcome by having duplicates of each reaction, and nominal correction is only used as a last resort. By manually comparing the overlapping nucleotide sequences and looking at the emissivity trace data as shown in Fig. 2-10, errors can be corrected and a complete overlap of fragments is obtained. The contiguous SSU rDNA sequence, which has a length of approximately 1000 bases, is saved as a consensus file for alignment against other known SSU rDNA sequences within the Genetic Data Environment package (GDE, Smith *et al.*, 1994).

2.6. Phylogenetic Reconstructions: Determining Evolutionary History

2.6.1. Nucleotide sequence alignment within the Genetic Data Environment Package

By comparing the rDNA sequences, it is possible to create foraminiferal phylogenies to evaluate the evolutionary history between, and within, taxa. The first stage of this process is to align the rDNA sequences so that the sequence data has positional homology (Swofford and Olsen, 1990). This means that at a given nucleotide position, each of the rDNA sequences have a nucleotide that is traceable to a common ancestor for all the taxa (Swofford and Olsen, 1990). The SSU rDNA sequences are manually aligned within GDE, which allows direct comparison of base position differences between the planktic foraminiferal species. Each foraminiferal genotype has different sequence lengths for the ~1000bp amplified region, varying between approximately 970-1200 bp, due to differing numbers of insertions and deletions. This requires gaps to be inserted into the sequences so that regions of positional homology may be built-up (Fig. 2-12).

The conserved regions of the rDNA sequences are alignable, but the variable regions and foraminiferal specific insertions (see Introduction for an explanation) show substantial sequence heterogeneity between, and sometimes within, taxa. These regions of the SSU rRNA gene are less alignable.

```
Type IIa        1  --GCA--CCA CAAG--AGCG T---GGAGTA -TGTGGCTTA ATTCGACTCA
Type IIb        1  --GCA--CCA CAAG--AGCG T---GGAGTA -TGTGGCTTA ATTCGACTCA
Type I          1  --GCA--CCA CAAG--AGCG T---GGAGTA -TGTGGCTTA ATTCGACTCA
Align-sites     1  ~~mmm--mmm mmmm--mmmm m---mmmmmm -mmmmmmmmmm mmmmmmmmmm

Type IIa       43  ACGCGCA-AC AATTTACTTG G------TCC GAACGC-TTT GGG--G-TTG
Type IIb       41  ACGCGCA-AC AATTTACTTG G------TCC GAACGC-TTT GAG--GATTG
Type I         41  ACGCGCA-AT AACTTACTTG G------TCC GAACGC-TTT GAG--GATTG
Align-sites    43  mmmmmmmm-mm mmmmmmmmmm m------mmm mmmmmm-mmm mmm--mmmmm

Type IIa       82  -ACAGTTATT G-------TA TAGTTCTGAT ATGAGAGGTC TTTGTAGTCA
Type IIb       81  -ACAGTTATT G-------TA TAGTTCTGAT ATGGGTGGTA TTTGTAGTCA
Type I         81  -ACAGTTTCT G-------GT ACAATGTGCG GCTTGTTTGT TACAACTACG
Align-sites    83  -mmmmmmmmm m--------- ---------- ---------- ----------

Type IIa      124  ACGTGTAGGT AGTTGTA--- ---------- ---------- ----------
Type IIb      123  ACGTGTAGGT AGTTGTA--- ---------- ---------- ----------
Type I        123  AATGATTCGA ATTGTTGTAA ATATGACTTG GCCGGCCTTC GGGTGTCTTG
Align-sites    93  ---------- ---------- ---------- ---------- ----------

Type IIa      141  ---------- ---------- ---------- ----TAT-TA AATATGAAA-
Type IIb      140  ---------- ---------- ---------- ----TAT-TA AATATGAAA-
Type I        173  GATCGGCGTA TCAGGTCGTA CATTG----- ----TAT-TC AATGAGAAA-
Align-sites    93  ---------- ---------- ---------- ----mmm-mm mmmmmmmmmm-

Type IIa      155  GT-------- ---------- -TCTTTTATG A------TTA TGTGATAG--
Type IIb      154  GT-------- ---------- -TCTTTTATG A------TTA TGTGATAG--
Type I        212  GT-------- ---------- -TCTTTTATG A------TTA TGTGGTAG--
Align-sites   107  mm-------- ---------- -mmmmmmmmm m------mmm mmmmmmmmm--

Type IIa      178  ---GTGGTG- CATGG-CCGT C-CTTAATTC GTGGA-GTGA TT-TGTC--T
Type IIb      177  ---GTGGTG- CATGG-CCGT C-CTTAATTC GTGGA-GTGA TT-TGTC--T
Type I        235  ---GTGGTG- CATGG-CCGT C-TTTAATTC GTGGA-GTGA TC-TGTC--T
Align-sites   130  ---mmmmmm- mmmmm-mmmm m-mmmmmmmm mmmmm-mmmm mm-mmmm--m

Type IIa      218  GCTT-AATTG CGCATT---- ---GCAAATT GTAATTGATC TTTTACCAGT
Type IIb      217  GCTT-AATTG CGCATT---- ---GCAAATT GTAATTGATC TTTTACCAGT
Type I        275  GCTT-AATTG CGCATT---- ---GCAAATT GTAATTGATC TTTGAACAGA
Align-sites   170  mmmm-mmmmm mmmmmm---- ---mmmmmmm mmmmmmmmmm mmmmmmmmm-
```

Fig. 2-12. An alignment of *T. quinqueloba* genotypes Types I, IIa and IIb to highlight the gaps inserted in the sequences to create positional homology. The aligned sites (m) are the bases used for phylogenetic analysis.

There are a number of databases available in which rDNA sequences are aligned, such as The Ribosomal Database Project (Maidak *et al.*, 1994, 1996). The alignment that I have used to import my North Atlantic planktic foraminiferal nucleotide sequences into, is a subset from an alignment created by Dr. Chris Wade (Division of Genetics, QMC,

University of Nottingham), who kindly provided me with this copy. From this I have added my rDNA sequences, and aligned them according to the unambiguously aligned sites decided by Dr. Wade. In addition to my own sequences I have imported a number of planktic foraminiferal sequences belonging to other workers which are published in the public domain of GenBank (http://www.ncbi.nlm.nih.gov/), and also a number of unpublished sequences kindly provided by Dr. Kate Darling (Department of Geology and Geophysics, University of Edinburgh).

The other sequences included in the phylogeny are: *G. bulloides* sequences from the Coral Sea (Darling *et al.*, 1997, GenBank Accession number U80793), the Southern Californian Bight (Darling *et al.*, 1999, unpublished sequence), the subantarctic (Darling *et al.*, submitted, unpublished sequence), and the Mediterranean Sea (de Vargas *et al.*, 1997, GenBank Accession number Z83957); *G. falconensis* from the Coral Sea (K. Darling, unpublished); *T. quinqueloba* from the subantarctic and the Coral Sea (Darling *et al.*, submitted, unpublished sequence); *G. ruber* from the Caribbean (Darling *et al.*, 1997, GenBank Accession number U65634), Coral Sea (Darling *et al.*, 1997, GenBank Accession number U80789), Southern Californian Bight (Darling *et al.*, 1999, GenBank Accession number AF102230), Caribbean (Pawlowski *et al.*, 1997, GenBank Accession number Z69599); *G. conglobatus* from the Coral Sea (Darling *et al.*, 1997, GenBank Accession number U80790) and the Caribbean (Pawlowski *et al.*, 1997, GenBank Accession number Z69600); *G. siphonifera* Type I from the Caribbean (Darling *et al.*, 1997, GenBank Accession number U65631; de Vargas *et al.*, 1997, GenBank Accession number Z83959); *G. siphonifera* Type II from the Caribbean (Darling *et al.*, 1997,

GenBank Accession number U80787), Coral Sea (Darling *et al.*, 1997, GenBank Accession number U80788), S. Californian Bight (Darling *et al.*, 1999, GenBank accession numbers AF102227 and AF102228); *Orbulina universa* from the Caribbean (Darling *et al.*, 1997, GenBank Accession number U65632), Coral Sea (Darling *et al.*, 1997, GenBank Accession number U80791), S. Californian Bight (Darling *et al.*, 1999, GenBank Accession number AF102229), Mediterranean Sea (de Vargas *et al.*, 1997, GenBank Accession number Z83961-2); *Globigerinoides sacculifer* from the Caribbean, and the Coral Sea (Darling *et al.*, 1997, GenBank Accession numbers U65633 and U80792); *N. pachyderma*, dextral (subantarctic, Darling *et al.*, submitted, unpublished sequence); *N. dutertrei* (Caribbean, Darling *et al.*, 1997, GenBank Accession number U65635); *G. glutinata* (Coral Sea, Darling *et al.*, 1999, unpublished sequence); and 15 benthic foraminifera (Pawlowski *et al.*, 1996, GenBank Accession numbers: *Allogromia* sp. X86093; *Ammonia beccarii* X86094; *Archaias angulatus* Z69603; *Peneroplis pertusus* Z69604; *Quinqueloculina* sp. Z69605; *Massilina secans* Z69606; *Astrammina rara* Z69608; *Astrorhiza triangularis* Z69609; *Textularia* sp. Z69610; *Bigerina* sp. Z69611; *Trochammina hadai* Z69612; *Bolivina* sp. Z69613; *Glabratella opercularis* Z69614; *Haynesina germanica* Z69615; *Elphidium aculeatum* Z69618).

For phylogenetic reconstruction, the sites/gaps that are out-with the homologous positions are cut out, so that only the alignable sites are remaining. These nucleotide positions are known as unambiguously aligned sites, and include all positions that can be attributed to ancestral positions common to all the taxa within the alignment. As the number of taxa increase within an alignment, the number of unambiguously aligned sites

decreases. For example, in the alignment with all of the sequences used in this study there are 505 unambiguously aligned sites from the ~1000 bp alignment. In contrast, an alignment containing only *Turborotalita quinqueloba* has as many as 761 unambiguously aligned sites. Therefore, when comparing genotypes from a single morphospecies the number of unambiguously aligned sites increases substantially.

Once the alignment has been cut down to only the unambiguously aligned sites the file can be output as a *.phylip file for phylogenetic analysis.

2.6.2. Distance based phylogenetic analysis

Distance based phylogenetic tree construction methods work on the principle of measuring the sequence dissimilarity between two sequences since they last shared a common ancestor. There are two components: (1) calculate a matrix of distances between sequences with correction (see below), according to a model; (2) trees are calculated from this distance matrix by an algorithm. A number of different tree construction methods exist, for example the Neighbour-joining (Saitou and Nei, 1987) or the Fitch-Margoliash (1967) method. Darling *et al.* (1997) found that the topology of the phylogenetic trees produced by both methods were relatively consistent, and that the neighbour-joining phylogeny was most similar to the phylogeny inferred from the fossil record. Further, the neighbour-joining method is highly reliable, and computationally fast relative to other reconstruction methods. In accordance with this, all distance based phylogenetic analyses were performed used the Neighbour-joining (Saitou and Nei, 1987)

(program NEIGHBOR) method within version 3.52c of the phylogeny inference package PHYLIP (Felsenstein, 1993).

The phylogeny inference package PHYLIP (Felsenstein, 1993) contains the programs necessary for phylogenetic reconstructions, e.g. DNADIST, SEQBOOT, NEIGHBOR, and CONSENSE. The steps undertaken to construct a neighbour-joining phylogeny are described in the following section.

2.6.2.1. Pairwise sequence comparison

After alignment, the sequences are output as a *.PHYLIP file which can then be used within the PHYLIP package. When calculating the sequence distances in a pairwise comparison, the observed distance tends to underestimate the actual number of differences between two sequences (Fig. 2-13). This is due to multiple substitutions occurring at the same nucleotide site over time, which therefore go unseen. This requires the distances calculated to be corrected, so they are not subject to such high inaccuracies as the uncorrected (raw) distances.

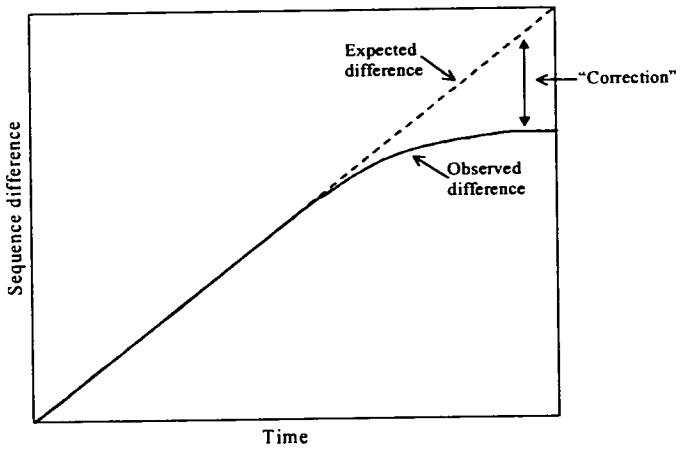


Fig. 2-13. Diagrammatic explanation of why distance calculations need to be corrected. The unseen (multiple) substitutions at a nucleotide site cause the observed difference between two sequences to tail-off, whereas the correction allows the expected sequence difference to be calculated (after Page and Holmes, 1998).

The model used to correct the sequence distances is the generalised 2-parameter (maximum likelihood F84) model (Felsenstein, 1993), within the program DNADIST, which also accounts for unequal rates of transition and transversion substitutions and unequal frequencies of the 4 nucleotides (Kishino and Hasegawa, 1989). A transition substitution is from a purine to a purine (A – G); or a pyrimidine to a pyrimidine (C – T). A transversion substitution is from a purine to a pyrimidine, or vice versa (e.g. A – C, or T – G). Transition substitutions are typically more frequent than transversion substitutions (for review see Page and Holmes, 1998). This procedure converts the aligned sequences to a pairwise distance matrix (Fig. 2-14), in which the evolutionary distances between each of the aligned sequences are shown. The evolutionary distances are calculated out of a total of one.

	Type I	Type IIa Caribb.	Type IIa N.Atlan.	Type IIa Coral Sea	Type IIa Califonia	Type IIb California
Type I						
Type IIa (Caribbean)	0.0718					
Type IIa (N. Atlantic)	0.0689	0.0026				
Type IIa (Coral Sea)	0.0704	0.0052	0.0026			
Type IIa (California)	0.0689	0.0026	0.0000	0.0026		
Type IIb (California)	0.0705	0.0225	0.0198	0.0199	0.0198	
Type IIb (N. Atlantic)	0.0705	0.0225	0.0198	0.0199	0.0198	0.0000

Fig. 2-14. Pairwise distance matrix for *Globigerinella siphonifera*, based on a 767 bp alignment. To calculate the evolutionary distance between two sequences, simply read along the horizontal line from one of the specific sequences to the column that corresponds to below the other specific sequence. This figure is the evolutionary distance between the two sequences under question.

The evolutionary distances can be read from the matrix directly. For example, the evolutionary distance between the Type I genotype and the Type IIa (Caribbean) genotype is 7.18 %. Further, the evolutionary distance between the Type IIa (North Atlantic) genotype and the Type IIa (California) genotype is 0.00 %, indicating that the SSU rDNA sequences are identical throughout the region used for analysis (767 bp).

2.6.2.2. Construction of the neighbour-joining tree

The distance matrix is then input into the program NEIGHBOR (Saitou and Nei, 1987), which calculates the neighbour-joining phylogeny based on the evolutionary distances between each of the sequences. A key feature of the neighbour-joining method is that it does not assume a molecular clock, such that lineage to lineage variations in evolution rate (branch length variations) are permitted. This allows two sequences which have a greater distance than a third sequence to cluster together, if that is the true

evolutionary relationship. As a molecular clock is not assumed, the neighbour-joining method compensates for this as follows. The matrix of estimated divergences between all pairs of sequences is used to calculate the net divergences from all other sequences. On the basis of this calculation, the separation between each pair of sequences (nodes) is adjusted, and a modified rate corrected distance matrix is constructed. This normalises the divergence of each sequence for its average clock rate, such that the molecular clock assumption is avoided.

The tree is constructed by joining the least distant pair of nodes from the modified matrix. When the two nodes are linked, their common ancestor is added to the tree and the terminal nodes are removed. This pruning process converts the newly added ancestor to a terminal node on a tree of reduced size. This process is repeated until only two nodes remain, separated by a single branch (see review by Swofford *et al.*, 1996).

A TREEFILE is produced which contains the data required to build a phylogenetic tree and this can be loaded into the TREETOOL package, which creates the neighbour-joining tree on the basis of the data in the treefile. The layout of the neighbour-joining tree can be improved by manipulation within treetool. Although the layout of the tree may change, evolutionary relationships between the sequences cannot be altered.

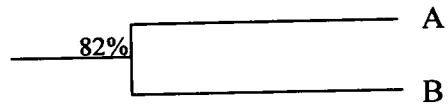
2.6.2.3. Bootstrap resampling

It is very important to know how significant the associations are between the genotypes within the neighbour-joining phylogeny. To assign statistical support to the neighbour-joining phylogeny, bootstrap resampling (1000 replicates) (Felsenstein, 1985)

was performed (programs SEQBOOT and CONSENSE). This resampling technique estimates the variance of the sampling distribution by repeatedly resampling the original data set. Data points from the original data set are sampled randomly, with replacement, until a new data set is generated. Some of the data points will be sampled once or more, and others will not be sampled at all (Swofford *et al.*, 1996). Therefore, only if all sites say the same thing will the same bootstrap tree be produced. If not, then each of the sample sets will thus show some differences in the bootstrap trees. This determines how frequently a particular association/branching pattern occurs.

The replicated data are input through the program DNADIST. This is the same as the pairwise sequence comparison described previously (section 2.6.2.1), with the exception that the distances are calculated for the number of replicates (1000). The distance file is then input into program NEIGHBOR to calculate the neighbour-joining phylogeny based on the evolutionary distances between each of the sequences, this time with 1000 repetitions. As in the initial use of NEIGHBOR, a treefile is produced which is input into the program CONSENSE. This program constructs a tree from the 1000 replicates, and shows the frequency at which particular associations occur. This produces a CONSENSUS TREE which illustrates the bootstrap support values for each fork in the phylogeny. According to Hillis and Bull (1993) a bootstrap support of 70 % or higher is considered to be statistically significant. For example, if a particular association has a bootstrap support of 82 %, then the association existed in 820 of the 1000 replicates within the phylogeny. In other words, in 82 % of the bootstrap replicates the branches to A and B fall to the right of that fork, hence the association between A and B has strong

bootstrap support.



The bootstrap values are then assigned to the neighbour-joining phylogeny produced earlier, and the neighbour-joining tree can then be manipulated to show the relationships as best as possible. To complete the phylogeny, the data from the distance matrix is used to accompany the evolutionary relationships between the genotypes on the neighbour-joining molecular phylogenetic tree.

2.7. Minimising contamination

Contamination whilst amplifying DNA is a major problem in molecular biology. Even a single molecule of contaminant DNA could be preferentially amplified producing unreliable PCR results. To minimise the possible risk of contamination between my own samples and between colleagues, the arrangement of the laboratories and the protocols used were designed specifically to address this problem.

To reduce contamination between co-workers, separate laboratory materials and equipment were used. This includes the gilson micro-pipettes used in all stages of the molecular work, the chemical reagents used in the various processes, and the electrophoresis tanks. By working in separate areas of the laboratory there is negligible

chance of cross contamination between samples obtained from different areas around the globe.

In all of the laboratories, 0.1 M hydrochloric acid is available which has the ability to breakdown DNA on contact. During the setting up of PCRs, the pipette tips and empty microcentrifuge tubes are disposed of into a tub containing HCl. The tube racks are also sprayed with HCl after every PCR is set-up, and in the main laboratory the racks are soaked regularly in a large container of acid. By having clean work surfaces contamination is also reduced. They can be sprayed with acid or 70% ethanol. The pipette tips and microcentrifuge tubes are autoclaved so that they are sterile. When DNA is being cut from agarose gels the possibility of contamination risk is high as residual DNA may be left behind on the illuminator. To reduce this risk a sheet of plastic food wrap is placed between the U.V. illuminator and the agarose gel so that DNA does not come into direct contact with the illuminator. When cutting is complete, the gel is disposed for incineration and the illuminator is wiped with 0.1 M HCl.

Chapter 3: The Foraminiferal Molecular Phylogeny

Index

3.1. Introduction

The foraminiferal fossil record extends back to the early Cambrian, approximately 550 Ma ago (Culver, 1991). Interpretation of the fossil record indicates that planktic foraminifers evolved from benthic foraminifers during the Mid-Jurassic or possibly even earlier (Loeblich and Tappan, 1974; Caron and Homewood, 1983). More recently, planktic foraminifers have been reported from the early Jurassic, ~200 Ma (Görög, 1994). It has been proposed that planktic foraminifera evolved from a single benthic lineage from which all further planktic foraminiferal species radiated (Tappan and Loeblich, 1988; Norris, 1991; Olsson *et al.*, 1992). However, Brinkhuis and Zachariasse (1988) suggested that, in addition to the evolution of planktic foraminifers from benthic species during the early Jurassic, planktic species also evolved later, after the Cretaceous-Tertiary boundary.

Molecular phylogenetic analysis (Darling *et al.*, 1996b; Pawlowski *et al.*, 1996; Wade *et al.*, 1996) indicated that the foraminiferal lineage most likely originated prior to the radiation of the major eukaryotic lineages (Fig. 3-1), approximately 1000-1100 Ma ago (Sogin, 1991; Knoll, 1992). Recent investigation by Pawlowski *et al.* (1999) has shown that some extant foraminifers do not possess a calcitic or proteinaceous shell but are amoeboid in form. They proposed that the first, early foraminifers were also "naked", and therefore were not preserved within the fossil record. This would explain why the foraminiferal fossil record extends back only to the Cambrian (Culver, 1991), when the molecular phylogeny indicates a much earlier origin for the foraminiferal lineage.

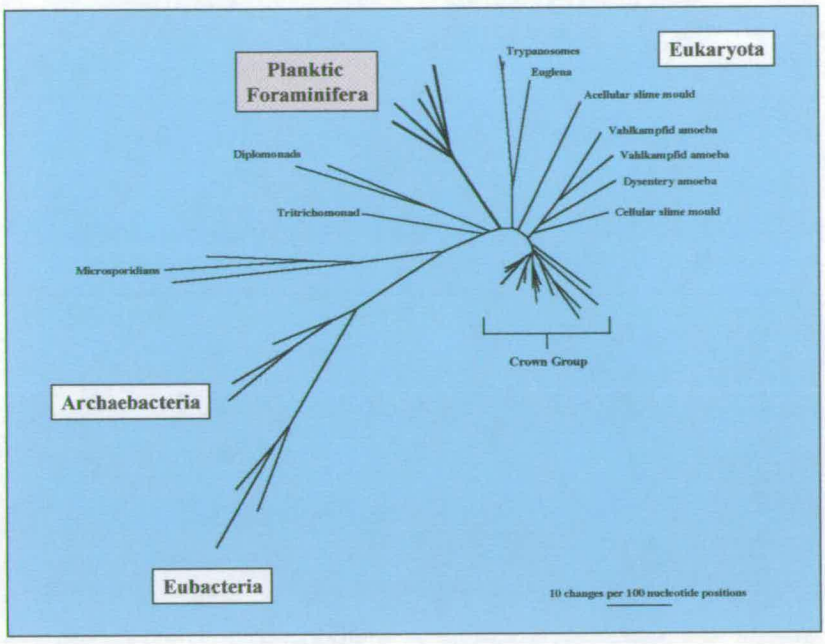


Fig. 3-1. Molecular phylogenetic "tree of life" (after Wade *et al.*, 1996), showing the early evolutionary placement of the planktic foraminifera prior to the major radiation of the crown group.

The planktic foraminifera form a monophyletic group within the "tree of life" (Fig. 3-1, Wade *et al.*, 1996). The structure of the group was examined in further detail by Darling *et al.* (1997) using a reconstructed molecular phylogeny containing both planktic (spinose and non-spinose) foraminifera and benthic foraminifera (Pawlowski *et al.*, 1997). It was found that the non-spinose planktic foraminifer *Neogloboquadrina dutertrei* clustered with the benthic foraminifers and separately from the spinose planktic foraminifers. This suggested that the planktic foraminifera are polyphyletic in origin (Darling *et al.* 1997), and has subsequently been confirmed by the addition of further taxa to the foraminiferal molecular phylogeny (Darling *et al.*, 1999; de Vargas *et al.*, 1997). Within the benthic/non-spinose region of the molecular phylogeny, the non-spinose planktic foraminifers do not all cluster together, indicating that they have evolved from separate benthic lineages.

In this chapter, a reconstructed foraminiferal molecular phylogeny is examined, which represents an extension of the phylogeny presented by Darling *et al.* (1999). The phylogeny has been extended to include 15 North Atlantic genotype sequences obtained during this study, 39 other published genotype sequences from GenBank, and also 7 unpublished genotype sequences provided by Dr. Kate Darling (see Chapter 2, section 2.6). The general topology of the foraminiferal molecular phylogeny is described, followed by a comparison of the branching pattern of the spinose planktic foraminifer genotype lineages with the interpretations of the fossil record. The chapter concludes with a summary of the main points regarding the foraminiferal molecular phylogeny.

3.2. Neighbour-joining molecular phylogeny

The methodology used in reconstructing the foraminiferal molecular phylogeny is described in Chapter 2 (section 2.6). A total of 46 planktic foraminiferal sequences, representing 7 genera, have been set against a background of 15 benthic foraminiferal sequences, which represent 5 benthic suborders. This permits an investigation of the evolutionary relationships between, and within, the planktic foraminiferal morphospecies.

A total of 15 genotypes, representing 225 specimens, obtained from the North Atlantic during this study were added to the foraminiferal phylogeny of Darling *et al.* (1999). The additional genotypes are members of the morphospecies clusters *Globigerina bulloides* d'Orbigny, *Globigerina falconensis* Blow, *Turborotalita quinqueloba* (Natland), *Neogloboquadrina pachyderma* (dextral) (Ehrenberg),

Globigerinita uvula (Ehrenberg), *Globigerinella siphonifera* (d'Orbigny), *Globigerinella calida* (Parker), and *Globigerinoides ruber* (d'Orbigny).

The main foraminiferal neighbour-joining phylogeny is constructed from 505 unambiguously aligned sites (Chapter 2, section 2.6). The whole alignment of both conserved and variable regions of the SSU rDNA fragment is shown in the appendices (A1.2). The 505 unambiguously aligned sites used for analysis were selected for their high degree of positional homology and are depicted as a series of m's (A1.2).

The reconstructed neighbour-joining foraminiferal molecular phylogeny is presented in Fig. 3-2 and the distance matrix is shown in the appendices (A1.3). The membraneous-walled benthic foraminifer *Allogromia*, which is thought to have diverged early in the history of the foraminiferal lineage (Tappan and Loeblich, 1988), has been used as a root for the tree. *Allogromia* provides an ancestral node, from which all other nodes are likely to have descended, allowing determination of evolutionary direction and ancestor-descendant relationships (Page and Holmes, 1998). The direction of evolution, within the molecular phylogeny (Fig. 3-2), is along the horizontal axis from left to right, i.e. from the benthic/non-spinose planktic foraminifers to the spinose planktic foraminifers. The vertical axis does not represent evolutionary direction.

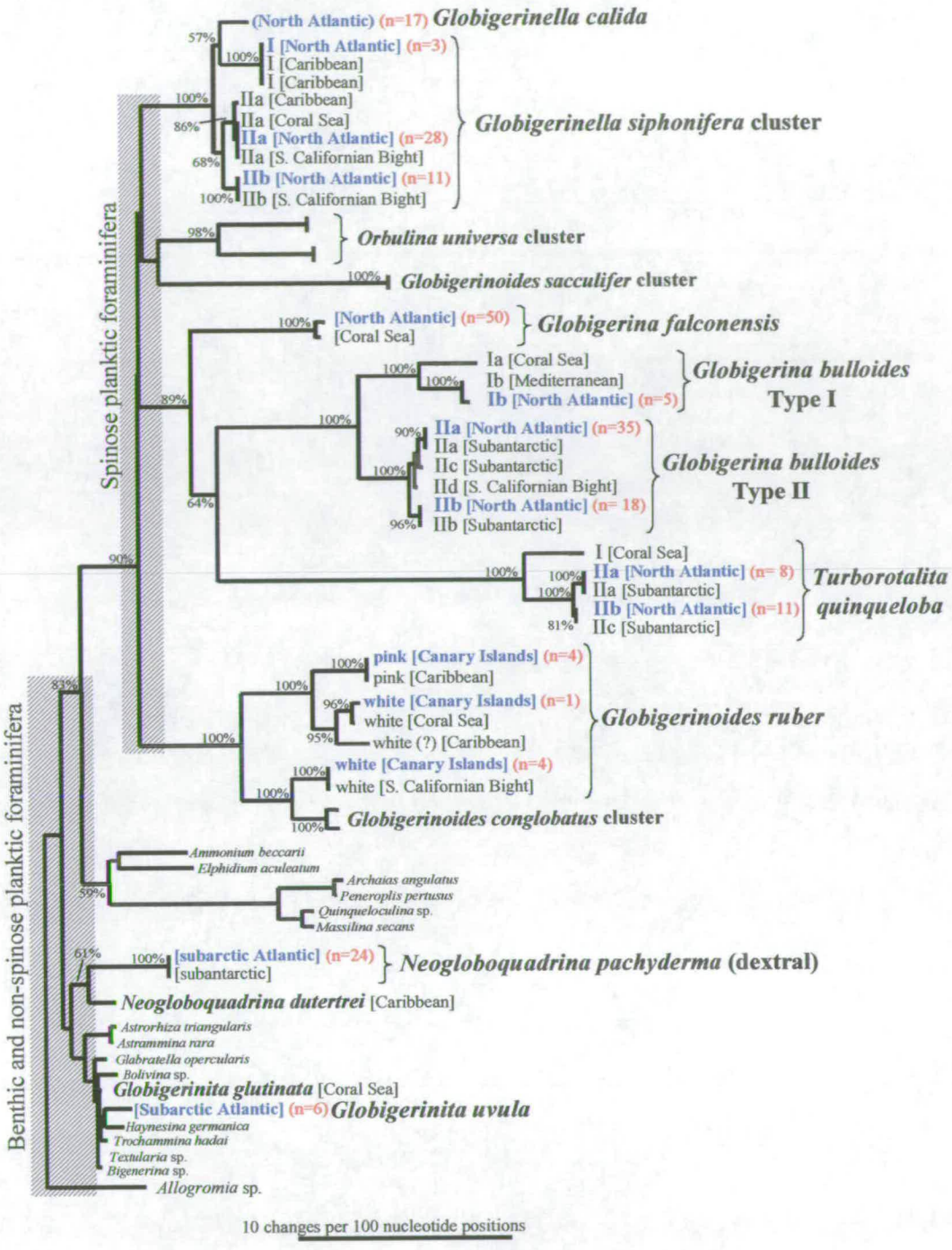


Fig. 3-2. Neighbour-joining molecular phylogeny based on 505 unambiguously aligned sites. Support has been assigned to the phylogeny using 1000 bootstrap replicates. The genotypes sequenced in this study are highlighted in blue, and the specimen numbers of each genotype are highlighted in red. More specific collection locations of the North Atlantic genotypes are shown in the individual chapters.

3.3. Topology of the foraminiferal molecular phylogeny

The neighbour-joining phylogeny (Fig. 3-2) is divided into two main groups (supported in 90 % of the bootstrap replicates), representing the division between the spinose planktic foraminifers and the benthic foraminifers/non-spinose planktic foraminifers. The long branch lengths of the planktic spinose group are indicative of high evolution rates (Darling *et al.*, 1997), since they only appeared in the fossil record approximately 27-30 Ma (Kennett and Srinivasan, 1983). On the other hand, the short branch lengths of the benthic taxa indicate extremely low evolution rates in the rRNA gene. Evolution in planktic foraminifers is though to be as much as 50-100 times faster than in some benthic foraminifers (Pawlowski *et al.*, 1997). Hillis *et al.* (1996) proposed a number of reasons behind variations in evolution rates, including the mutation effects of ultraviolet radiation on DNA during replication, differences in organismal generation time, differences in DNA replication in germ line cells, metabolic rate, nucleotide generation times, and DNA repair efficiency. This could perhaps account for the rate differences between the benthic foraminifers and the spinose planktic foraminifers. Within the spinose planktic foraminifer region of the phylogeny there is considerable variation in evolution rate, which is reflected in the branch length variations that exist between, and within, genotype clusters (Fig. 3-2). The highest evolution rate is observed in the *Globigerina bulloides* and *Turborotalita quinqueloba* lineages, and the lowest evolution rate is observed in the *Globigerinella* genotype cluster (Fig. 3-2). The differences in evolution rate between spinose planktic lineages may be due to differences in generation time and genome turnover, with high gamete numbers increasing the chance of replication based mutations (Darling *et al.* 1999).

The spinose planktic foraminiferal cluster is characterised by a series of monophyletic groups (Fig. 3-2). The order of divergences within the spinose planktic foraminiferal group is unresolved due to the short internodes and low bootstraps between the main monophyletic groups, indicating that the lineages most likely radiated over a short period of time. This is consistent with fossil record interpretations which suggest that the present day spinose planktic foraminiferal lineages radiated in the Oligocene, approximately 27-30 Ma, from an early *Globigerina* species (Kennett and Srinivasan, 1983). The ancestral *Globigerina* is thought to have been *Globigerina officinalis* (Blow, 1979; Spezzaferri and Premoli Silva, 1991).

The non-spinose planktic foraminifers *Neogloboquadrina pachyderma* and *Globigerinita uvula* cluster amongst the benthic foraminifers, consistent with the polyphyletic origins of the planktic foraminifers (Darling *et al.*, 1997). Within this region of the phylogeny, the branch lengths are relatively short indicating slow evolution rates. Due to the low evolution rate, and the short internodes between branches, the precise evolutionary relationships between the non-spinose foraminiferal species remain unresolved.

3.4. The spinose planktic foraminiferal region of the molecular phylogeny and comparison to the fossil record

The monophyletic groups of genotypes within the spinose planktic foraminiferal cluster represent morphologically defined species. There is substantial branch length variation between the groups, which is discussed below together with a comparison of the order of branching within the phylogeny against the fossil record.

3.4.1. *Globigerina/Turborotalita* lineage

This monophyletic group has relatively strong bootstrap support (89 %), and comprises three distinct lineages. The extant morphospecies representing these lineages are *Globigerina bulloides*, *Globigerina falconensis*, and *Turborotalita quinqueloba*. The large genetic distances between each morphospecies suggest a relatively ancient divergence, probably soon after the major radiation of the spinose species. The clusters have very different branch lengths, with *G. falconensis* having a much lower evolution rate than either *G. bulloides* or *T. quinqueloba*.

Interpretations of the fossil record suggest that *Globigerina praebulloides* first appeared ~ 30 Ma, during the late Oligocene (Pearson, 1993; Berggren *et al.*, 1995). This gave rise to *G. bulloides* during the middle Miocene, ~ 16 Ma (Kennett and Srinivasan, 1983). It has been suggested that *G. falconensis* diverged from the *Globigerina* lineage ~18 Ma (Kennett and Srinivasan, 1983). Interpretation of the fossil record suggests that the *T. quinqueloba* lineage diverged ~ 25 Ma, prior to the divergence of *G. falconensis* (Spezzaferri, 1994). In contrast, the molecular phylogeny (Fig. 3-2) suggests that all three lineages diverged over a relatively short period.

The reconstructed molecular phylogeny (Fig. 3-2) shows that the *G. bulloides* and *G. falconensis* genotypes cluster separately from one another, forming two well-defined monophyletic groups. Further, a large mean evolutionary distance (15.8 %) separates the two monophyletic groups. For an explanation of how this evolutionary distance was derived, please refer to the Appendix (A1.4). This is the procedure by which all evolutionary distances are calculated throughout this thesis. These findings confirm the conclusions of Malmgren and Kennett (1977) that *G. bulloides* and *G. falconensis* are distinct species. *Globigerina bulloides* and *G. falconensis* also differ

biologically since *G. falconensis* bears symbionts, and *G. bulloides* does not (Hemleben *et al.*, 1989). In addition, isotope analyses suggest that the two species have different habitats within the water column, with *G. falconensis* living nearer the surface than *G. bulloides* (Malmgren and Kennett, 1977).

3.4.2. *Globigerinella* lineage

This monophyletic group has strong bootstrap support (100 %), and is represented by two extant morphospecies: *Globigerinella siphonifera* and *Globigerinella calida*. Within the spinose group, the *Globigerinella* lineage has the shortest branch lengths, indicating the slowest evolution rate. Interpretations of the fossil record suggest that the ancestral species was *Globigerinella obesa*, which first appeared during the late Oligocene (~29 Ma) (Spezzaferri and Premoli Silva, 1991). The first appearance of *G. siphonifera* has been suggested at ~ 14 Ma, and for *G. calida* at ~ 4 Ma (Kennett and Srinivasan, 1983). There is considerable discrepancy between the molecular phylogeny and the fossil record, with regard to the first appearance of *G. calida*. The molecular phylogeny (Fig. 3-2) suggests that both *G. siphonifera* and *G. calida* have been in existence for a similar period of time, whereas interpretation of the fossil record suggests a much more recent appearance for *G. calida*. If a constant evolution rate is assumed across the entire *G. calida* lineage, between the spinose radiation (27-30 Ma) and the end of the *G. calida* branch (present day), a datum of 10-11 Ma is indicated.

3.4.3. *Globigerinoides ruber/conglobatus* lineage

This monophyletic group has strong bootstrap support (100 %). This cluster has previously been described by Darling *et al.* (1999), where a comparison between the molecular phylogeny and the fossil record is discussed in detail. The phylogeny shows that there are two extant *G. ruber* lineages, represented by two clusters which form well defined monophyletic groups (Fig. 3-2). One of the *G. ruber* clusters is more closely related to *G. conglobatus* than to the other *G. ruber* cluster. The common ancestor between the two lineages is probably *Globigerina primordius*, which diverged from *Globigerina praebulloides* at ~ 29 Ma (Spezzaferri and Premoli Silva, 1991). Darling *et al.* (1999) suggested that the two extant lineages diverged from the ancestral *Globigerinoides* species during the early Miocene, and that the lineages were represented by *Globigerinoides subquadratus* and *Globigerinoides obliquus*. The proposal that *G. ruber* was represented by these lineages was initially made by Cordey (1967).

3.5. The non-spinose planktic foraminifers

The non-spinose planktic foraminiferal morphospecies *N. pachyderma* and *G. uvula* cluster within the benthic and non-spinose planktic region of the 505 bp molecular phylogeny. The *Neogloboquadrina pachyderma* (dextral) genotypes cluster with *Neogloboquadrina dutertrei* (Fig. 3-2), separated by an evolutionary distance of 4.9 %. Their association is supported in 61 % of the bootstrap replicates. The microperforate planktic foraminiferal species *Globigerinita uvula* is separated by an evolutionary distance of 1.6 % from the microperforate planktic foraminifer

Globigerinita glutinata, although their precise placement within this region of the phylogenetic tree remains unresolved.

The evolutionary ancestry of both *N. pachyderma* and *G. uvula* is poorly understood. Indeed, there is contention as to when each of these morphospecies first appeared. The literature describing the first appearance of *N. pachyderma* is reviewed by Kennett and Srinivasan (1980), with dates ranging from the Late Miocene (Bandy, 1960, 1972; Cifelli, 1961, 1973; Berger, 1970; Parker and Berger, 1971; Kennett and Srinivasan, 1980, 1983) to the Pleistocene (Olsson, 1976). There is also some uncertainty when *G. uvula* first appeared. Saito *et al.* (1981) suggested that it has a stratigraphic distribution from the Pleistocene to Recent. In stark contrast however, Srinivasan and Kennett (1983) suggest a more ancient appearance, with a stratigraphic distribution from the Late Oligocene to Recent. As the evolutionary relationships within the non-spinose planktic/benthic region of the molecular phylogeny remain unresolved due to the slow evolution rate of the taxa in this region of the phylogeny, a detailed comparison with the fossil record cannot be carried out.

3.6. Summary

The molecular phylogeny of Darling *et al.* (1999) has been extended to include North Atlantic planktic foraminiferal genotypes. The additional 15 genotypes represent 8 morphospecies, across 6 genera. Consistent with recent foraminiferal molecular phylogenies, the spinose planktic foraminiferal genotypes cluster separately from the benthic and non-spinose planktic foraminiferal genotypes. Within the benthic and non-spinose planktic region of the phylogeny, the non-spinose foraminifers *N. pachyderma*

and *G. uvula* fall separately, indicating that they have evolved from different benthic ancestors.

In general, the molecular phylogeny matches previous interpretations of the fossil record. There are however, discrepancies regarding the first appearance of some morphospecies in the fossil record, and their evolutionary position within the molecular phylogeny. In particular, the molecular phylogeny suggests that the first appearance of *G. calida* is considerably more ancient than fossil record suggests, and that the *G. bulloides*, *G. falconensis* and *T. quinqueloba* lineages all diverged at approximately the same time. In addition to the fossil record interpretation being highly subjective, the discrepancies are most likely explained by "cryptic speciation", whereby molecular divergences have occurred prior to any morphological change. These divergences would then go unnoticed within the fossil record until morphological change took place.

It is clear that there are some difficulties when comparing the molecular phylogeny with the fossil record. When the molecular evolution rate differences between, and within, morphospecies are considered, it is evident that any attempt to calibrate the foraminiferal molecular phylogeny utilising fossil record interpretation is highly problematic.

Chapter 4: *Globigerina/Turborotalita* cluster

Index

4.4. The *Globigerina falconensis* cluster

4.4.1. Sequence homogeneity of *G. falconensis* and distribution within the North Atlantic

4.4.2. Morphological variability of *G. falconensis*

4.5. Discussion

4.5.1. Distribution of *Globigerina/Turborotalita* genotypes within the North Atlantic

4.5.1.1. Genotype distribution in the subarctic region

4.5.1.2. Genotype distribution in the transitional-subtropical regions

4.5.2. Relationship between genotype and the ocean environment

4.5.3. Morphology/genotype relationships

4.5.4. Cryptic speciation within *G. bulloides* and *T. quinqueloba*

4.5.5. Is *Globigerina falconensis* of palaeoceanographic importance ?

4.5.6. Gene flow

4.1. Introduction

Within this chapter, *Globigerina bulloides* d'Orbigny, *Globigerina falconensis* Blow and *Turborotalita quinqueloba* Natland are investigated. They are grouped and discussed together in the same chapter as molecular analysis shows that they form a monophyletic group within the planktic spinose region of the foraminiferal molecular phylogeny (Fig. 3-2). The group splits into three separate lineages, two of which are *Globigerina* and one *Turborotalita*. The extant morphospecies of these lineages are distributed across a large proportion of the world's oceans, from polar/subpolar to tropical water masses (Bé and Tolderlund, 1971; Bé, 1977; Malmgren and Kennett, 1977; Hemleben *et al.*, 1989). They represent an important component of the global marine sediment flux and their calcitic shells provide a very important source of information for palaeoceanographic and palaeoclimatic study.

The present day distribution of each morphospecies within the oceans is outlined. This is followed by a summary of relevant research to date for each of the morphospecies, and its significance for palaeoceanography is outlined. For each morphospecies, the evolutionary relationships between the genotypes are examined in detail. The distribution patterns of the genotypes found during the North Atlantic collections are illustrated. An investigation of obvious morphological variability, such as coiling direction and cytoplasm colouration, found within the morphospecies completes each section. The chapter is completed with an overview discussion of the results obtained from the *Globigerina/Turborotalita* morphospecies.

4.1.1. *Globigerina bulloides* d'Orbigny

Presently, *Globigerina bulloides* occurs predominantly in subpolar water masses, but it is also common in upwelling areas and boundary currents in low latitude regions (Bé and Tolderlund, 1971). It is found over a temperature range of 0°C to 27°C, with highest abundance between 3°C to 19°C (Bé and Tolderlund, 1971). In subpolar regions it frequently exceeds 50 % of the total population (Bé and Tolderlund, 1971). Boltovskoy *et al.* (1996) compared South Atlantic planktic foraminiferal distributions to North Atlantic distributions (Bé and Hamlin, 1967; Ottens, 1991). It was shown that *G. bulloides* is almost twice as abundant in the North Atlantic compared with the South Atlantic. In the Northeastern Atlantic, *G. bulloides* is found predominantly in the upper 60 m of the water column (Schiebel *et al.*, 1997).

In addition to the general distribution pattern, there is also considerable seasonal and spatial variation in the abundance of *G. bulloides* within more localised areas. This has been observed from studies using both plankton tows and sediment trap material (e.g. Deuser *et al.*, 1981; Sautter and Thunell, 1991a; Pujol and Grazzini, 1995; Ufkes *et al.*, 1998). For example, within the Sargasso Sea, Deuser *et al.* (1981) found that *G. bulloides* abundance peaked in winter and spring, with a sharp drop in abundance in summer. This is in general agreement with the seasonal distribution pattern observed by Pujol and Grazzini (1995) in the Mediterranean Sea, where *G. bulloides* accounts for 20-40% of the total planktic foraminiferal assemblage in winter along the North African coast. In contrast, however, was the high abundance (50% of the total assemblage) during the summer in the far west of the Mediterranean Sea (Pujol and Grazzini, 1995). Further, in the Southern Californian Bight, *G. bulloides* was the dominant species from January to the end of March, and also during the

upwelling period (late April to mid-June) (Sautter and Thunell, 1991a). They showed that *G. bulloides* is abundant across a wide temperature range, and that its abundance may increase due to upwelling (with the associated high phytoplankton production) and/or during the spring bloom.

Ortiz *et al.* (1995) suggested that food limitation was important in the distribution of foraminifera, such as *G. bulloides*, and that it had a food threshold, below which it could not survive. Since *G. bulloides* is asymbiotic (Hemleben *et al.*, 1989), Ortiz *et al.* (1995) suggested that the food threshold effect may be more pronounced. In addition, they suggested that *G. bulloides* is likely to be omnivorous since in addition to being more abundant during high phytoplankton conditions, it has been observed to ingest copepods.

As *G. bulloides* occurs across a wide geographic range, test morphology has been examined between areas. Bé (1977) noted that subpolar and subtropical/tropical forms of *G. bulloides* are very similar in test morphology, although he proposed that the subpolar form differed subtly from the lower latitude form in having a more deeply recessed aperture, a less pronounced apertural rim, and a test width that almost equals the length of the test. Further, the mean test size of *G. bulloides* was found to increase with decreasing sea-surface temperature (Malmgren and Kennett, 1976), and that the percentage of sinistral individuals was a function of temperature, with a higher percentage of sinistral individuals in cooler water (Boltovskoy, 1973; Malmgren and Kennett, 1976).

The morphological variations within *G. bulloides* have been considered useful for palaeoceanographic reconstructions, both as an indicator of cold climates and upwelling within the fossil record (Malmgren and Kennett, 1978; Kroon, 1991; Naidu

and Malmgren, 1996a, 1996b). As *G. bulloides* forms such a large part of the planktic foraminiferal flux, it provides a substantial source of information for palaeoceanographers utilising transfer functions and also geochemical data (Bard *et al.*, 1987; Clemens *et al.*, 1991; Sautter and Thunell, 1991b; Anderson and Prell, 1993; Kroon *et al.*, 1997; Bemis *et al.*, 1998; Ganssen and Kroon, in press). The geochemical investigations have largely been confined to $\delta^{18}O$ and $\delta^{13}C$ ratios, however more recently Cd/Ca and Mg/Ca records are also being utilised (e.g. Elderfield *et al.*, 1998; Rickaby and Elderfield, 1999) to estimate palaeoproductivity and palaeotemperature.

4.1.2. *Turborotalita quinqueloba* (Natland)

Turborotalita quinqueloba bears symbionts (Hemleben *et al.*, 1989) and predominantly inhabits subpolar waters but is also found in polar and transitional waters (Bé and Tolderlund, 1971). It is found across a temperature range of 1-21°C, with a preference for temperatures of less than 12°C (Bé and Tolderlund, 1971). It is commonly found together with *G. bulloides*, and is therefore a major component of cold water masses (Bé and Tolderlund, 1971). The frequency of *T. quinqueloba* generally decreases from high to low latitudes (Parker, 1962). The highest abundance of *T. quinqueloba* within the Pacific occurs during phytoplankton blooms (especially diatom blooms), and periods of upwelling (Sautter and Thunell, 1991a), most likely in response to preferred light and, or, nutrient conditions.

In examining a north-south transect of sediment core-top assemblages, Parker (1962) and Parker and Berger (1971) noted morphological variability within *T. quinqueloba*. Further investigation of morphological variability within *T. quinqueloba* was completed by Kroon *et al.* (1988) with biometric analyses on living populations

collected from the eastern North Atlantic, western Mediterranean, Red Sea and northern Indian Ocean. It was shown that the diameter of the last whorl, final chamber and penultimate chamber correlate closely to sea surface temperature, but that other morphological characteristics such as lobateness and aperture type appear to be less correlated to sea surface temperature. The relation of these growth characteristics to salinity and density shows different trends for the eastern North Atlantic-Mediterranean surface waters, compared to the Indian Ocean. This led Kroon *et al.* (1988) to suggest that this was either due to a single species possessing phenotypes that are sensitive to temperature, rather than salinity and density, or that there is a North Atlantic and a northern Indian Ocean subspecies/genotype. The study also examined morphological and isotopic variations within *T. quinqueloba* from an eastern Mediterranean piston core, and showed that size parameters were strongly correlated with the isotopic signal, possibly reflecting the environmental conditions of the photic zone of the water column.

A further study, by Bauch (1994), examined test size variations of *T. quinqueloba* from sediment cores obtained from the Norwegian-Greenland Sea. It was shown that the mean test size of *T. quinqueloba* increased after the Last Glacial Maximum, and that there was an east-west decrease towards smaller test sizes. At present in the Norwegian-Greenland Sea, the sea surface temperature and salinity also decrease in a westerly direction (Dietrich, 1969; Swift, 1986). However, Bauch (1994) could not identify the particular characteristic that could account for this observed trend in test size, whether it was temperature, salinity, nutrients or some other factor. He concluded that test size variations within *T. quinqueloba* may prove to be a valuable tool for palaeoceanographic reconstructions. It has become apparent that for *T.*

quinqueloba to be utilised in full, the smaller size fractions (63-150 μm) of the sediments must be examined in addition to the larger fractions (Carstens et al., 1997; Bauch, 1994). This is highlighted by the case of Fram Strait (Arctic) sediments (Carstens et al., 1997). They showed that if only the large fractions (>150 μm and >200 μm) were considered the assemblage consisted almost entirely of *N. pachyderma*. However, if the small fraction was examined the assemblage comprised of 2 species (60 % *N. pachyderma* and 38 % *T. quinqueloba*), instead of a monospecific assemblage. This has considerable implications for palaeoenvironmental reconstructions utilising the large fraction of marine sediments, and suggests that *T. quinqueloba* may prove a more useful proxy in the future.

4.1.3. *Globigerina falconensis* Blow

At present, *G. falconensis* is found in tropical to temperate water masses, with peak abundance in water >13°C (Saito *et al.*, 1981). *Globigerina bulloides* and *G. falconensis* are morphologically very similar (Hemleben *et al.*, 1989), and in some cases, workers have combined the two forms together in distributional studies. Initial non-biometric work on the *G. bulloides-G. falconensis* plexus by Bé (1969) and Kennett (1969) suggested that the two forms showed morphological gradation, hence were thought to represent subspecies or phenotypic variants. However, Malmgren and Kennett (1977) examined these two forms and found them to differ in all essential morphological aspects, concluding that they are distinct species. The two species clearly differ in biology since, in contrast to *G. bulloides*, *G. falconensis* bears symbionts (Hemleben *et al.*, 1989). In addition to having different optimum distributional patterns, isotope data indicates that they also have different depth

habitats, with *G. falconensis* living nearer the surface than *G. bulloides* (Malmgren and Kennett, 1977). Apart from inclusion within transfer function based investigations, *G. falconensis* has not specifically been used in palaeoceanography.

4.2. Molecular relationships within the *Globigerina bulloides* cluster

A total of 44 *G. bulloides* specimens was collected from the subarctic Atlantic (Fig. 2-2) for phylogenetic analysis, and a total of 14 *G. bulloides* specimens was collected from the transitional-subtropical North Atlantic during Poseidon 247 and Meteor 37/2 cruises (see Figs. 2-1 and 2-3). The *G. bulloides* cluster is supported in 100 % of bootstrap replicates within the 505 bp phylogeny (Fig. 4-1).

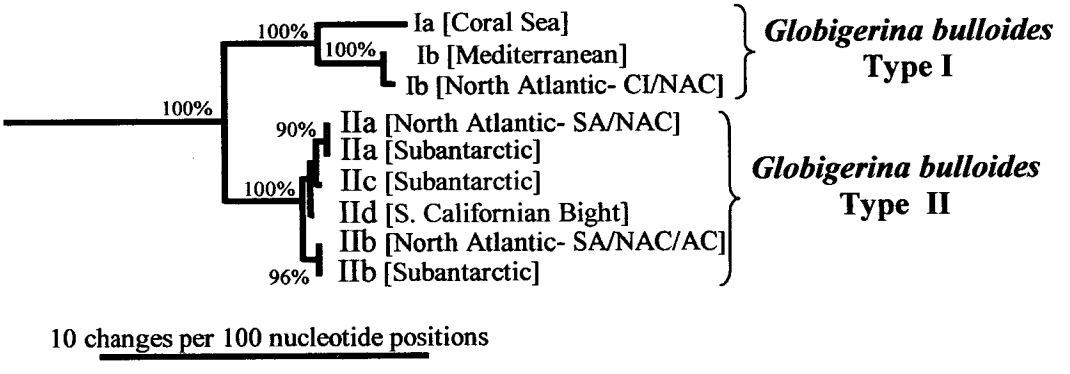


Fig. 4-1. Subset of neighbour-joining phylogeny based on 505 unambiguously aligned sites (Fig. 3-2). Included are the genotypes from the *Globigerina bulloides* morphospecies. Included with the North Atlantic genotypes are genotypes from the Southern Californian Bight (Darling *et al.*, 1999) and the Subantarctic (Darling *et al.*, submitted), and the Mediterranean Sea (de Vargas *et al.*, 1997). Support has been assigned to the phylogeny using 1000 bootstrap replicates. The region of the North Atlantic from which the genotypes were obtained is shown. **CI** denotes the Canary Islands, **SA** the subarctic, **NAC** the North Atlantic Current, and **AC** the Azores Current.

The *G. bulloides* lineage divides into 2 distinct clusters (Type I and Type II), which are each supported in 100 % of the bootstrap replicates respectively. The Type I and Type II clusters are separated by a mean evolutionary distance of 8.4 %. Within the subarctic region of the North Atlantic two genotypes (Types IIa and IIb) have been identified and within the transitional-subtropical North Atlantic three genotypes (Types Ib, IIa and IIb) have been identified

4.2.1. The *G. bulloides* Type I cluster

Within the 505 bp phylogeny (Fig. 4-1) the Type I cluster has 100 % support from the bootstrap replicates. The branch further divides representing the division between the North Atlantic (Canary Islands/NAC) / Mediterranean genotype and the Coral Sea genotype (Fig. 4-2). The Coral Sea genotype (Type Ia) is quite divergent from the North Atlantic genotype (Type Ib) with an evolutionary distance of 4.7 %. The North Atlantic genotype clusters with the Mediterranean genotype (de Vargas *et al.*, 1997, GenBank Accession number Z83957). This association is supported from 100 % of the bootstrap replicates. The North Atlantic Type Ib genotype is practically identical to the Mediterranean *G. bulloides* genotype. However, the differences that do exist between the North Atlantic and Mediterranean genotypes occupy positions which could be accounted for by errors from manual sequencing by de Vargas *et al.* (1997).

4.2.2. The *G. bulloides* Type II cluster

Within the 505 bp phylogeny (Fig. 4-1) the Type II cluster has support from 100 % of the bootstrap replicates. The Type II cluster includes genotypes from the

North Atlantic: Type IIa (subarctic and NAC) and Type IIb (subarctic/NAC/AC). Also included in the Type II cluster are genotypes from the subantarctic Atlantic (Types IIa, IIb and IIc, Darling *et al.*, submitted), and the Southern Californian Bight (Type IId, Darling *et al.*, 1999). The evolutionary relationships within the Type II cluster are not completely resolved within the 505bp phylogeny due to the low number of base differences within the conserved regions used for phylogenetic analysis. To attempt to resolve the evolutionary relationships within this cluster a within morphospecies phylogeny was reconstructed, substantially increasing the number of bases that could be unambiguously aligned. The phylogeny shown in Fig. 4-2 is based on 851 unambiguously aligned sites, and illustrates the close evolutionary relationships that exist between the genotypes from the Type II cluster. The alignment and distance matrix is shown in the Appendix (A2.1.1 and A2.2.1). The *G. bulloides* Type Ia genotype (Coral Sea, Darling *et al.*, 1997) has been used as an outgroup, and illustrates how divergent the Type I genotypes are to the Type II genotypes.

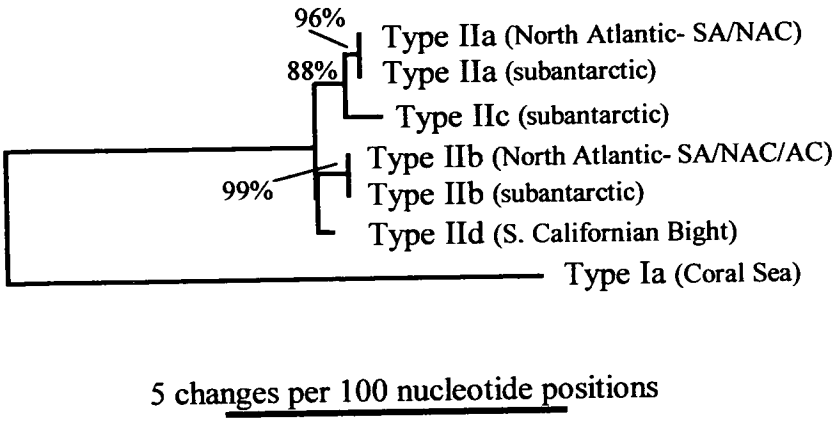


Fig 4-2. Neighbour-joining phylogeny of *Globigerina bulloides* genotypes based on 851 unambiguously aligned sites. Support has been assigned to the phylogeny using 1000 bootstrap replicates. The phylogeny used the Coral Sea Type Ia genotype as an outgroup. The region of the North Atlantic from which the genotypes were obtained is shown. **SA** denotes the subarctic, **NAC** the North Atlantic Current, and **AC** the Azores Current.

The 851 bp phylogeny (Fig. 4-2) has improved the resolution of the Type II cluster. It is clear that *G. bulloides* Type IIa and Type IIc cluster together, with their association having relatively high bootstrap support (88 %), but the relationship between Type IIb and Type IId remain unresolved. In an attempt to resolve all the evolutionary relationships within the Type II cluster, a phylogeny was constructed only using the Type II genotypes to maximise the number of base positions for analysis. The phylogeny was based on 935 unambiguously aligned sites, and is presented in Fig. 4-3. The alignment and distance matrix is shown in the Appendix (A2.1.2 and A2.2.2).

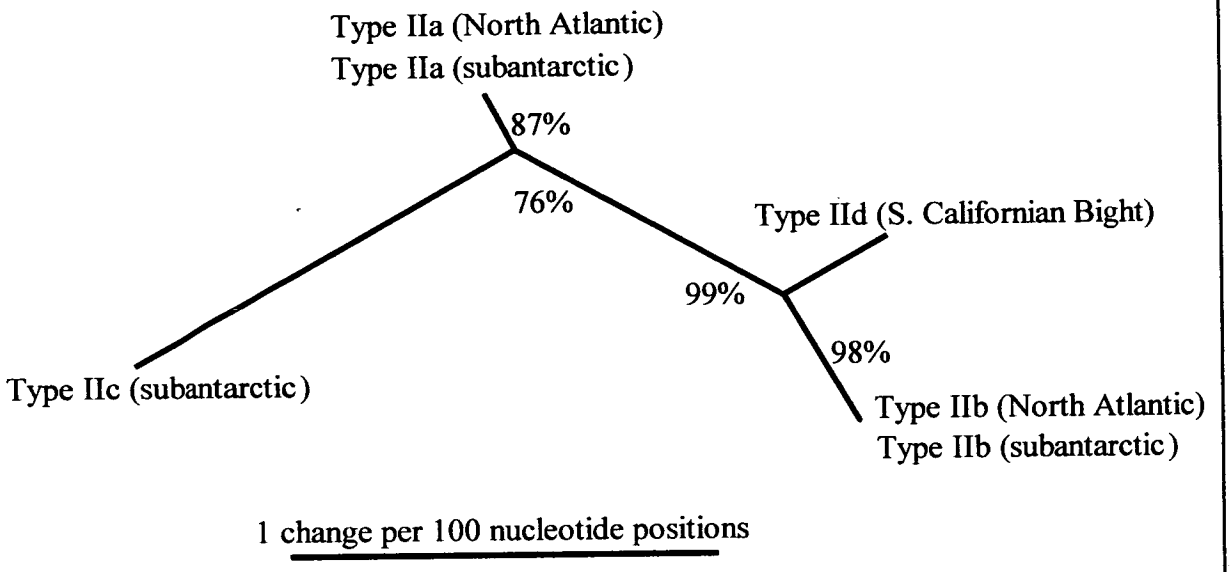


Fig. 4-3. Neighbour-joining phylogeny of *Globigerina bulloides* Type II genotypes based on 935 unambiguously aligned sites. Support has been assigned to the phylogeny using 1000 bootstrap replicates. The phylogeny highlights the evolutionary relationships that exist between genotypes of the *Globigerina bulloides* Type II cluster.

The North Atlantic *G. bulloides* Types IIa and IIb are separated by a genetic distance of 1.2 % within the 935 bp phylogeny (Fig. 4-3). Identical genotypes (Type IIa and IIb) were obtained from the subantarctic Atlantic by Darling *et al.* (1999). Those bipolar genotypes were identical within the conserved regions used for phylogenetic analysis, and also throughout the variable regions of the ~1000bp SSU rDNA fragment. The 935 bp phylogeny (Fig. 4-3) shows that Type IIa and Type IIc cluster together, and that Type IIb and Type IId cluster together. These associations are supported in 76 % and 99 % of the bootstrap replicates respectively. Within the 935 bp phylogeny, Type IIa is separated from each of the other genotypes by an evolutionary distance of 1.2 %, and the Type IIb genotype is separated from Type IIc and Type IId by evolutionary distances of 2.2 % and 0.6 % respectively. The evolutionary distance between two genotypes can vary when calculated from

phylogenies based on different numbers of unambiguously aligned sites. Therefore, to avoid confusion, comparison of evolutionary distances between genotypes must be restricted to the phylogeny in question.

The variable regions of the ~1000 bp region of the SSU rDNA fragment provides further information regarding the level of divergence between the genotypes. By comparing aligned sequences, the number of base differences between pairs of genotypes can be determined. These differences can be divided up into base substitutions and sequence length variations (base deletions/insertions). An illustration of the terms base substitution and sequence length variation is shown in Fig. 4-4. A base substitution refers to a specific nucleotide position that differs in base composition between two sequences (see Chapter 2, section 2.6.2.1). An insertion or deletion refers to an alignable region between sequences that differ in length due to the presence of gaps in one or more of the sequences.

a. Sequence length variations

Type IIa	CGTCGT	GTTAG	TTCTTCC
Type IIb	CGTCGT	GTT--	TTCTTCC

b. Base substitutions

Type IIa	TGTCGGGT	ATGTGGCC	TCGGTCAT
Type IIb	TGTCGGGT	GCGTGGCC	TCGGTCGT

Fig. 4-4. A short alignment of *G. bulloides* Types IIa and IIb illustrating sequence length variations (base deletions/insertions) (a), and base substitutions (b).

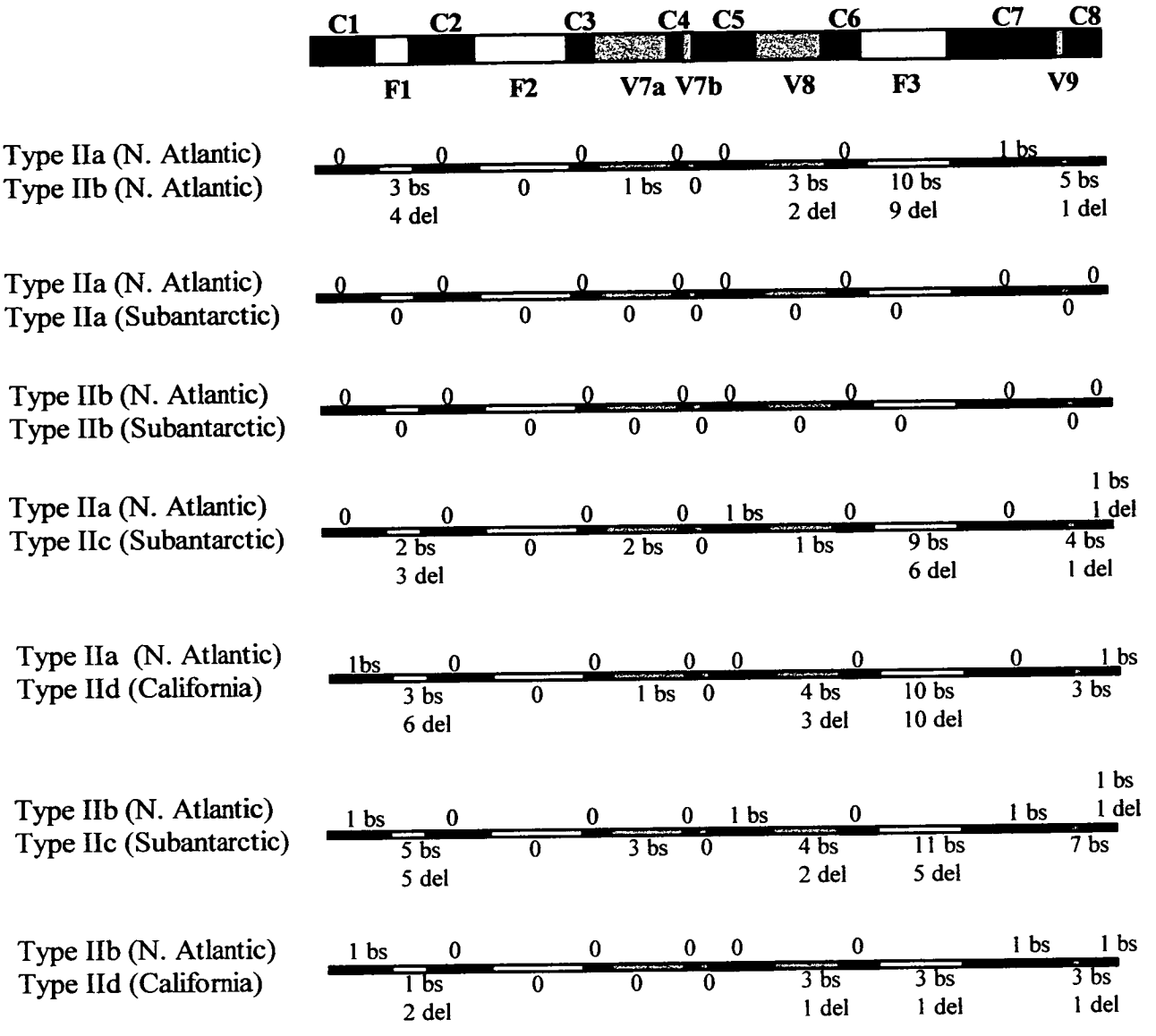


Fig. 4-5. Schematic representation of the ~1000bp 3' terminal region of the SSU rRNA gene. Comparisons of the genetic differences are made between the genotypes from the North Atlantic, Southern Californian Bight, subantarctic Atlantic, and Coral Sea. Indicated are the number of base differences between each of the genotypes, with the number of base substitutions and sequence length variations (base deletions/insertions) shown. **C1-C8** represent the highly conserved regions which were aligned relative to comparable regions present in all eukaryotes, **V7-V9** represent variable length expansion segments present in most eukaryotes and **F1-F3** represent three insertions which are unique to foraminifera (after Darling *et al.*, 1997). **bs** represents base substitutions and **del** represents sequence length variations due to insertions/deletions.

A within-genotype analysis of their ~ 1000 bp fragment is presented in Fig. 4-5. Comparison of the Type IIa and Type IIb genotypes shows that a total of 23 base substitutions and 16 base insertion/deletion differences are observed across the ~1000 bp SSU rDNA sequence. The Type IIa and Type IIb genotypes from the North Atlantic and the subantarctic show complete sequence homogeneity throughout the entire ~1000 bp SSU rDNA sequence. A total of 20 base substitutions and 11 insertion/deletion differences exist between the Type IIa genotype and the Type IIc (subantarctic) genotype. Comparison of the Type IIa and Type IId genotypes (S. Californian Bight) shows that they differ by 23 base substitutions and 19 insertion/deletion differences. Comparison of the Type IIb and Type IIc genotypes shows that they differ by 34 base substitutions and 13 insertion/deletion differences. The Type IIb and Type IId genotypes differ by 13 base substitutions and 5 insertion/deletion differences.

Consistently, within the genotype comparisons (Fig. 4-5) it is apparent that the foraminiferal specific insertion, F3, has the highest number of base differences. Conversely, in all of the genotype comparisons the foraminiferal specific insertion, F2, has no base differences (Fig. 4-5).

4.2.3. Distribution of *G. bulloides* genotypes within the subarctic Atlantic

A total of 44 *G. bulloides* SSU rDNA partial sequences were obtained from the subarctic collection. The subpolar *G. bulloides* specimens comprised Type IIa and Type IIb genotypes, with 32 specimens and 12 specimens being sequenced respectively. The distribution of *G. bulloides* Types IIa and IIb throughout the subarctic transect is shown in Fig. 4-6.

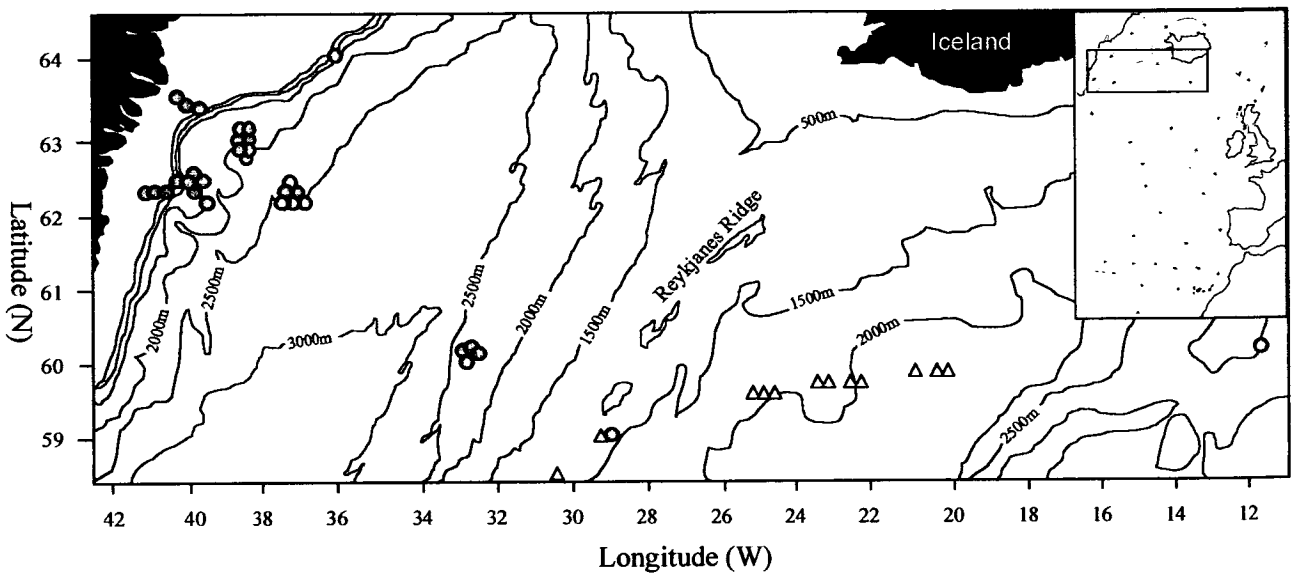


Fig. 4-6. The distribution of *G. bulloides* Type IIa (○) and IIb (△)genotypes within the subarctic Atlantic. The inset map indicates the region expanded in the main part of the figure, and the major surface current systems within the eastern North Atlantic (see Chapter 1).

A total of 11 specimens of *G. bulloides* was sequenced from the transect east of the Reykjanes Ridge. All of the specimens were of the Type IIb genotype, apart from a single Type IIa individual which was found at the far east of the transect (Fig. 4-6). Three *G. bulloides* specimens were sequenced from the waters above the Reykjanes Ridge, which were found to comprise one Type IIa specimen and two Type IIb specimens. From the transect west of the Reykjanes Ridge 30 specimens of *G. bulloides* were sequenced, all of which were found to be the Type IIa genotype. No Type IIb genotype was found west of the Reykjanes Ridge (Fig. 4-6). The transect distribution is suggestive of two distinct populations, a Type IIa population and a predominantly Type IIb population, with a mixed zone above the Reykjanes Ridge.

4.2.4. Distribution of *G. bulloides* genotypes within the transitional/subtropical assemblage zones of the North Atlantic

A total of 14 *G. bulloides* SSU rDNA partial sequences were obtained from the transitional/subtropical assemblage zones during the Poseidon 247 and Meteor 37 collections. Within these regions *G. bulloides* was generally scarce, although its highest numbers were found in the transitional waters of the North Atlantic Current (NAC). The distribution of *G. bulloides* genotypes within the transitional/subtropical assemblage zones of the North Atlantic is illustrated in Fig. 4-7.

A total of 3 distinct *G. bulloides* genotypes were obtained from these regions of the North Atlantic: Type Ib, Type IIa and Type IIb. Within the NAC, all three genotypes were found to co-exist within the water column (Fig. 4-7).

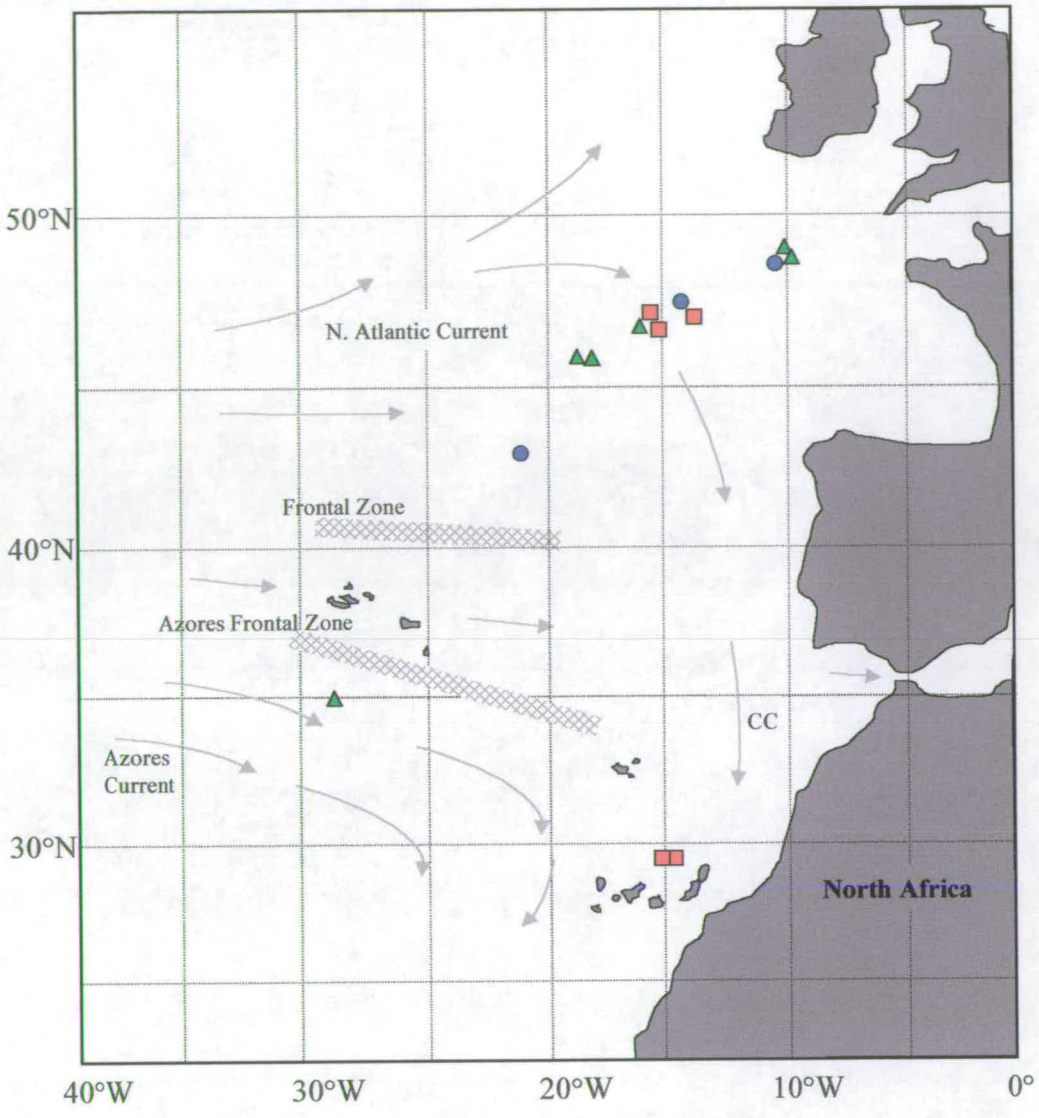


Fig. 4-7. The distribution of *G. bulloides* genotypes within the North Atlantic, as found during collections M38/2 and P247. The genotypes are denoted: Type Ib (■), Type IIa (●), and Type IIb (▲). The major surface currents are indicated. **CC** represents the cool Canary Current. The approximate location of the main watermass frontal zones are shown, as determined by the shipboard thermo-salinometer.

Three specimens of *Globigerina bulloides* Type Ib were obtained from the NAC (12.5°C-12.9°C) and a further two specimens were obtained from the waters north of the Canary Islands (~20°C). Three specimens of *Globigerina bulloides* Type IIa were found within the NAC (12.1-12.9°C), with one of those specimens obtained as far south as 43°N (Fig. 4-7) where the water was 14.2°C. A total of five specimens of

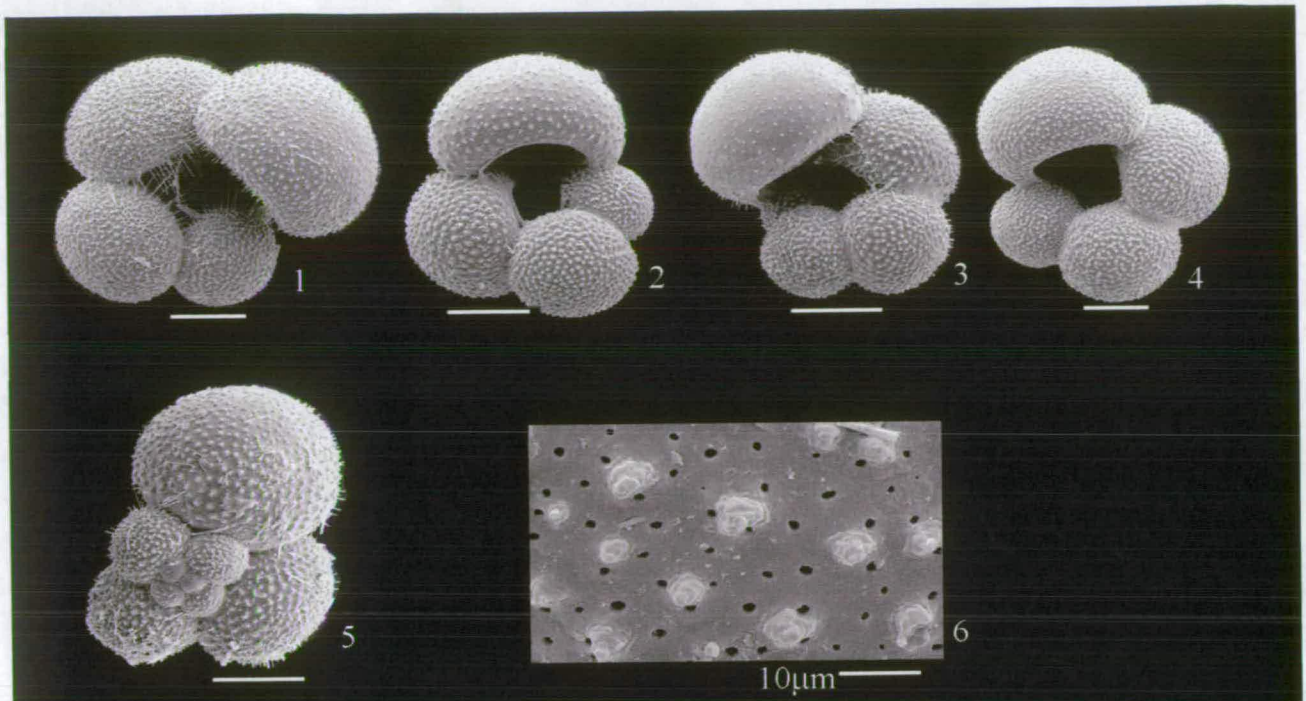
Globigerina bulloides Type IIb was obtained from the transitional waters of the NAC (12.1-13.4°C). One further specimen of the *G. bulloides* Type IIb genotype was obtained from the subtropical waters of the Azores Current at 35°N (Fig. 4-7) where the water was 19.3°C.

4.2.5. Morphological variability of *G. bulloides* specimens

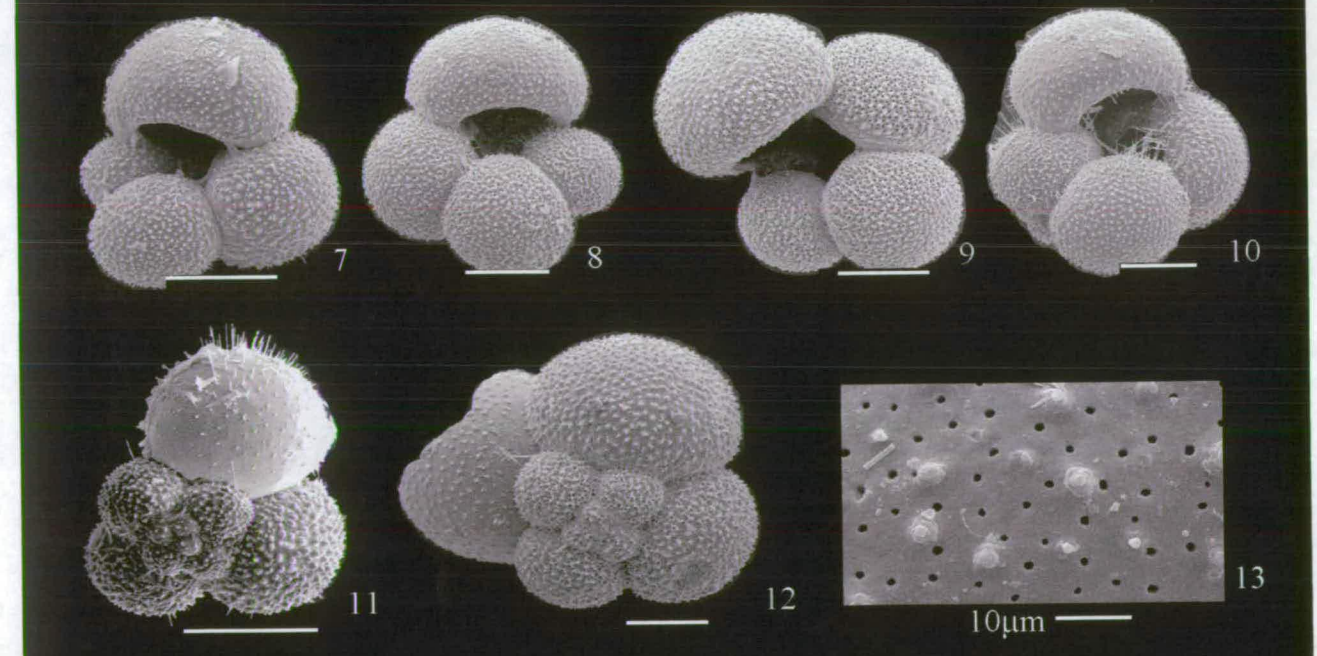
a. Subarctic specimens

One of the main problems in morphological studies using plankton collected from pumping surface water, or by net tows, is that most of the specimens obtained are juvenile and have not reached their mature size or form. This hinders the investigation of morphological difference between specimens, as comparisons can only be made against specimens within the sediments once they have reached their mature state.

Subarctic *G. bulloides* displayed a range of morphological variability (Plate 4-1). The specimens shown have not all reached their mature size and form. The genotypic investigation of the subarctic *G. bulloides* (section 4.2.3) indicated the presence of a region with a Type IIa genotype population, and a region with a predominantly Type IIb genotype population (Fig. 4-6). The specimens are therefore split into two separate groups reflecting the morphotypes found in each region.



Globigerina bulloides Type IIa morphotypes



Globigerina bulloides Type IIb morphotypes

Plate 4-1

Plate 4-1 description: *Globigerina bulloides* d'Orbigny from the subarctic Atlantic. The specimens are divided into Type IIa and Type IIb representatives. The Type IIa representatives were collected west of the Reykjanes Ridge, and the Type IIb representatives were collected from east of the Reykjanes Ridge (Fig. 4-6). The scale bars represent 100μm, unless otherwise indicated.

Type IIa (west of the Reykjanes Ridge):

Figs. 1-2: sinistrally coiling specimens, umbilical side. **Figs. 3-4:** dextrally coiling specimens, umbilical side. **Fig. 5:** sinistrally coiling specimen, trochospiral side. Note the well defined juvenile chambers, indicating its ontogenic development. **Fig. 6:** Enlargement of the terminal chamber of the specimen in Fig. 1. The test pores and the spine bases are visible. Note the small pore sizes compared to the larger pores observed in *G. ruber* (see Plate 6-1).

Type IIb (east of the Reykjanes Ridge):

Figs. 7, 9 & 10: dextrally coiling specimens, umbilical side. **Fig. 8:** sinistrally coiling specimen, umbilical side. **Fig. 11:** sinistrally coiling specimen, trochospiral side, juvenile specimen. Note the smooth, newly formed last chamber. **Fig. 12:** sinistral coiling specimen, trochospiral side, mature specimen. **Fig. 13:** Enlargement of the final chamber of the specimen in Fig. 10. The test pores and the spine bases are visible.

The percentage of sinistrally coiling individuals at each bulk plankton collection locality is shown in Fig. 4-8. It is clear that there are consistently more sinistral specimens, since within the geographic range of the Type IIa genotype the average coiling direction was 66 % sinistral, and within the geographic range of Type IIb, 58 % (Fig. 4-8). However, it is evident from Fig. 4-8 that there is no real difference in coiling direction between the two populations. Indeed the error bars suggest that the average coiling direction may be similar for both populations.

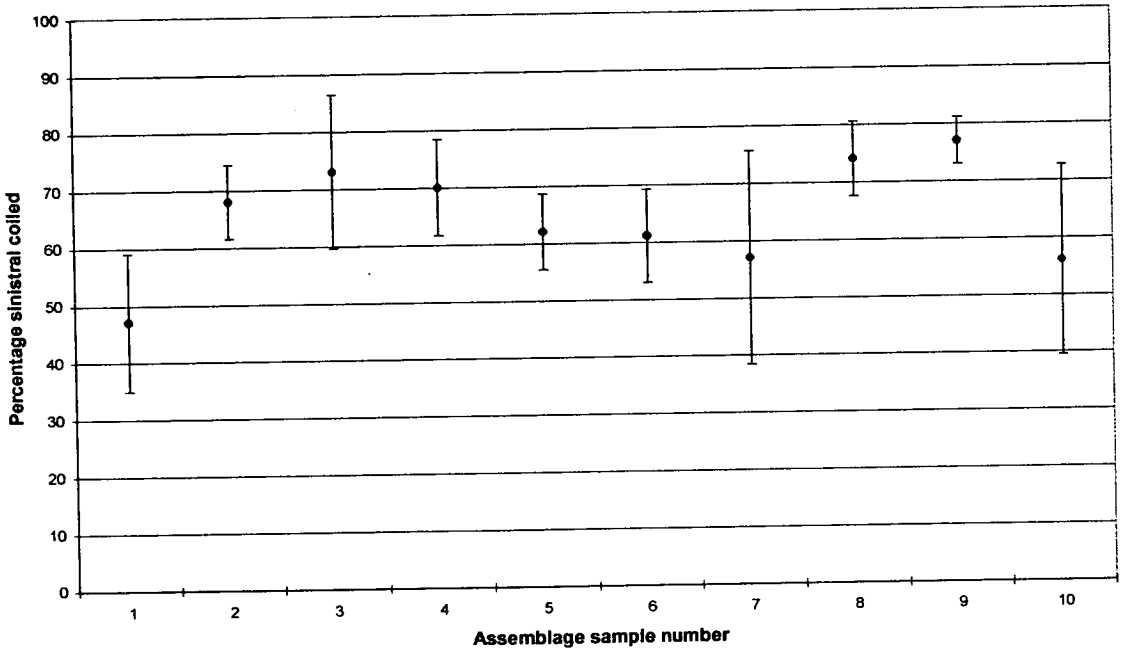


Fig. 4-8. *G. bulloides* coiling directions for each ashed bulk plankton assemblage (for details see Table 2-3). Assemblages 1 and 2 are from east of Reykjanes Ridge, and assemblages 3 to 10 are from west of Reykjanes Ridge. The mean percentage of sinistrally coiled specimens for each assemblage is shown. In addition, the standard deviation from the mean for each assemblage is shown as an error bar. This takes into account the variation in specimen numbers obtained in each sample.

Cytoplasmic colour was noted for all specimens used for DNA analysis. Orange was the predominant colour (75 %), though red, brown and yellow were also recorded. There was no correlation between genotype and cytoplasmic colour, nor between geographic location and cytoplasmic colour.

b. Transitional/subtropical specimens

The number of *G. bulloides* specimens obtained from ashed bulk plankton samples was too low to provide any data regarding coiling direction in these regions. In addition, as three genotypes were found to co-exist within the North Atlantic Current they have not been differentiated morphologically. The specimens obtained from bulk plankton samples are shown in Plate 4-2.

Of the *G. bulloides* Type Ib, Type IIa, and Type IIb specimens sequenced from the transitional/subtropical North Atlantic, each genotype displayed sinistrally and dextrally coiled individuals. However, the numbers sequenced were too low to provide accurate data regarding coiling direction ratios. As found in the subarctic *G. bulloides* specimens, the predominant cytoplasmic colouration of transitional/subtropical *G. bulloides* was orange, with brown and yellow cytoplasm also being observed.

Interestingly, one of the *G. bulloides* Type Ib specimens sequenced had an apertural lip (specimen not shown). Additional investigation will be required to determine whether this is a common morphological feature of this genotype, but it has the potential to confuse identification between this genotype and *G. falconensis*.

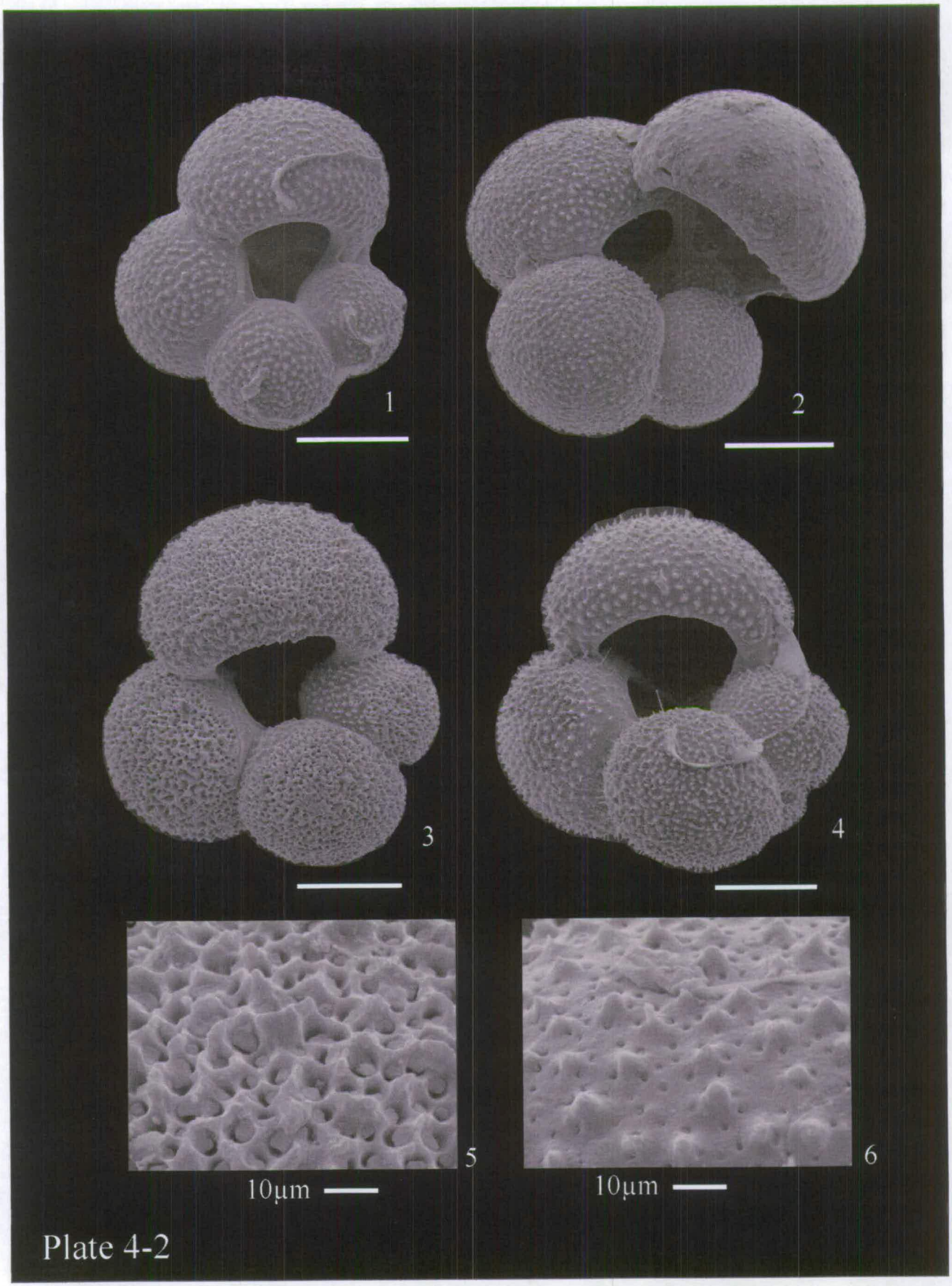


Plate 4-2

Plate 4-2 description: *Globigerina bulloides* d'Orbigny from the transitional-subtropical Atlantic. The specimens were obtained from the NAC, the NATW, and the AC (Fig. 4-7). The scale bars represent 100μm, unless otherwise indicated.

Fig. 1. NAC specimen, umbilical side. **Fig. 2.** NATW specimen, umbilical side. **Fig. 3.** NATW specimen, umbilical side, note heavy calcification on chambers compared with Figs. 1, 2 and 4. **Fig. 4.** AC specimen, umbilical side. **Fig. 5.** Enlargement of terminal chamber of same specimen shown in Fig. 3. Pore sizes relatively large compared with Fig. 6 due to heavier calcification. **Fig. 6.** Enlargement of terminal chamber of specimen shown in Fig. 4.

4.3. Molecular relationships within the *Turborotalita quinqueloba* cluster

A total of 19 *T. quinqueloba* specimens was collected from the subarctic Atlantic (see Fig. 2-2) for phylogenetic analysis. Within the 505 bp phylogeny (see Fig. 3-2) the *T. quinqueloba* cluster is supported in 100 % of the bootstrap replicates. The *T. quinqueloba* branch divides, representing the division between the subtropical genotype (Type I) and the subpolar genotypes (Type II). The Type I genotype is separated from the Type II genotypes by a mean evolutionary distance of 5.7 %. The Type II cluster is supported in 100 % of the bootstrap replicates. The Type I and Type II genotypes are also highly divergent within the variable regions of the SSU rDNA fragment. As this is cannot be illustrated using molecular phylogenetic analysis, the aligned sequences can be found in the Appendix (A2.1.3). There is also substantial sequence length variation throughout the partial SSU rDNA fragment. The distance matrix for this phylogeny is also shown in the Appendix (A2.2.3).

4.3.1. The *T. quinqueloba* Type II sub-cluster

Within the 505 bp phylogeny, the Type II cluster has support from 96 % of the bootstrap replicates (Fig. 3-2). As the precise evolutionary relationships are not resolved within the conservative 505 bp phylogeny, a within morphospecies phylogeny was constructed (Fig. 4-9) based on 762 unambiguously aligned nucleotide sites. The alignment and distance matrix is shown in the Appendix (A2.1.3).

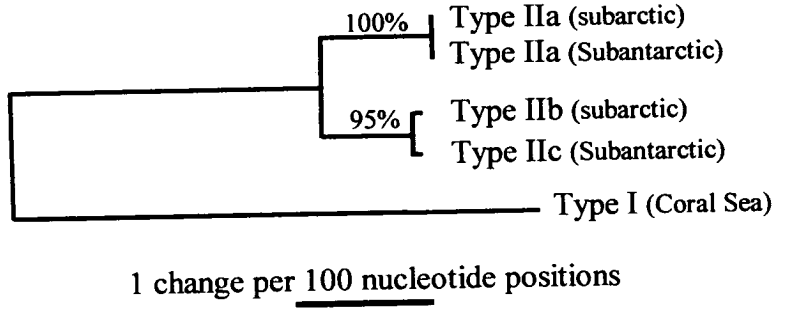


Fig. 4-9. Neighbour-joining phylogeny of *Turborotalita quinqueloba* genotypes based on 762 unambiguously aligned sites. Support has been assigned to the phylogeny using 1000 bootstrap replicates. The phylogeny used the Coral Sea Type Ia genotype as an outgroup. The evolutionary relationships between the *T. quinqueloba* Type II genotypes are shown.

Within the 762 bp phylogeny the *T. quinqueloba* Type II branch divides, representing the division between the Type IIa genotype cluster and the Type IIb and Type IIc genotype cluster (Fig. 4-9). The clusters are supported in 100 % and 95 % of the bootstrap replicates respectively. Two genotypes of *T. quinqueloba* (Types IIa and IIb) were identified within the subarctic Atlantic. Within the 762 bp phylogeny (Fig. 4-9), the clusters are separated by a mean evolutionary distance of 1.6 %. As the low evolutionary distance still significantly masks the actual number of sequence variations

that exist between the two genotypes, a detailed comparison of genotype base substitutions and sequence length variations is presented in Fig. 4-10.

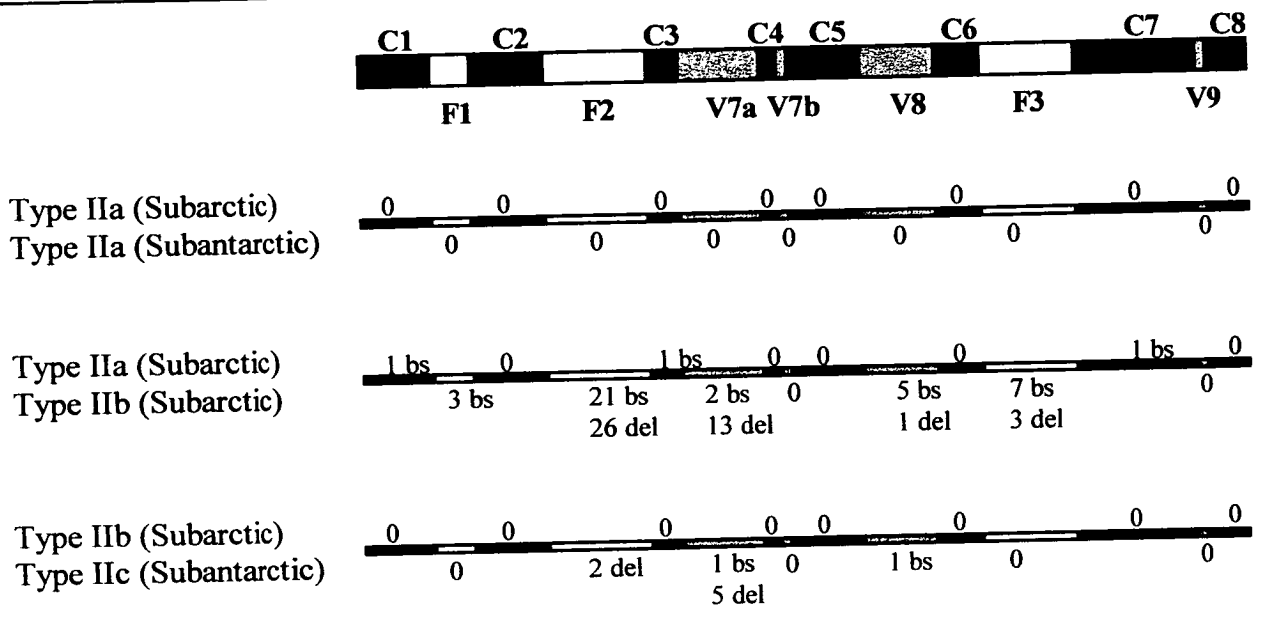


Fig. 4-10. Schematic representation of the 3' terminal region of the SSU rRNA gene (~1200 bp) comparing *Turborotalita quinqueloba* Type II genotypes from the subarctic and subantarctic Atlantic. The number of substitutional base changes, and sequence length variations (base insertions/deletions), that exist between the genotypes are shown. **C1-C8** represent the highly conserved regions which were aligned relative to comparable regions present in all eukaryotes, **V7-V9** represent variable length expansion segments present in most eukaryotes and **F1-F3** represent three insertions which are unique to foraminifera. **bs** represents base substitutions and **del** represents sequence length variations.

The Type IIa and Type IIb genotypes differ by 41 base substitutions and 43 insertion/deletion differences across the entire ~1200 bp region. In particular, the foraminiferal specific insertions, F2 and F3, and the variable length expansion segments, V7a and V8, have a large number of base differences between Types IIa and IIb (Fig. 4-10). The Type IIa genotype from the subantarctic (Darling *et al.*, submitted) was found to be identical to the subarctic Type IIa genotype throughout the entire ~1200 bp region of the SSU rDNA fragment (Fig. 4-10). The Type IIb (subarctic) and Type IIc (subantarctic, Darling *et al.*, submitted) genotypes are very closely related,

being separated by an evolutionary distance of only 0.1 % within the 762 bp phylogeny (Fig. 4-9). These genotypes have only 2 base substitution and 7 base insertion/deletion differences across the ~1200 bp SSU rDNA fragment (Fig. 4-10).

4.3.2. Distribution of *T. quinqueloba* genotypes within the subarctic Atlantic

A total of 17 *T. quinqueloba* genotypes was obtained from the subpolar North Atlantic. Of each genotype, 8 Type IIa and 9 Type IIb were sequenced respectively. In addition, two further specimens of the Type IIb genotype were obtained from the transitional waters north of Scotland (59°45.7'N/05°46.9'W). The distribution of *T. quinqueloba* Type IIa and Type IIb throughout the subpolar transect is shown in Fig. 4-11.

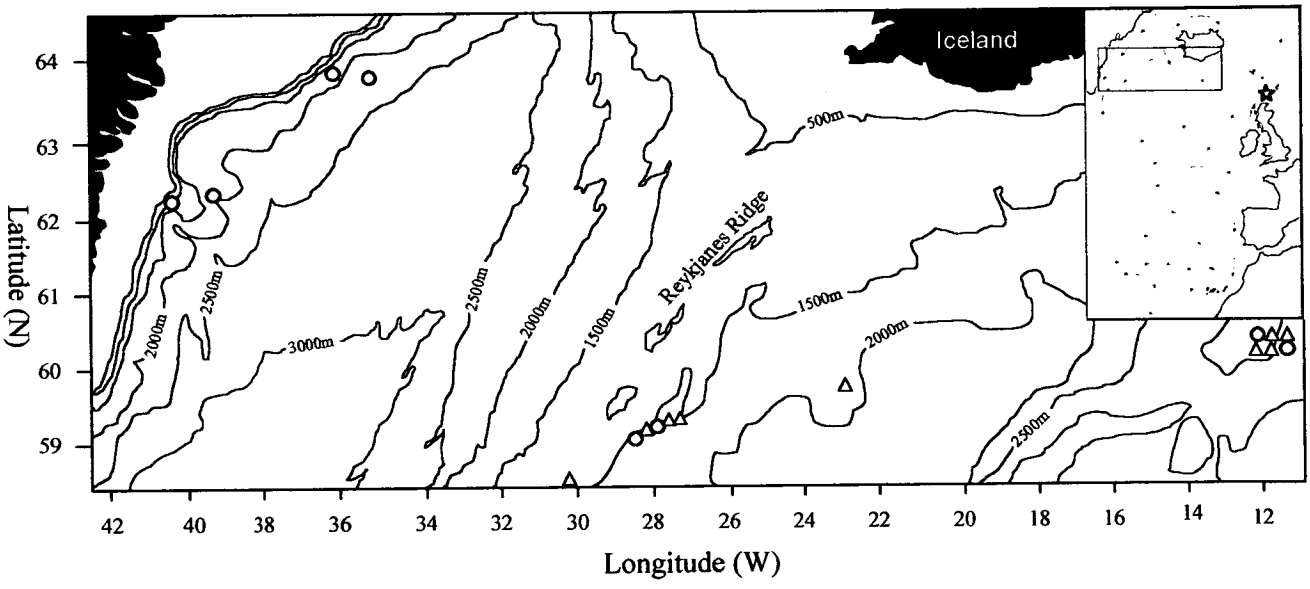


Fig. 4-11. Distribution of *T. quinqueloba* Type IIa (○) and IIb (△) genotypes within the subarctic Atlantic. The inset map indicates the location of two additional Type IIb genotypes ☆, and the major surface current systems of the region.

Both genotypes co-exist in the water column within the eastern section of the subpolar transect. However, only Type IIa specimens were found west of the

Reykjanes Ridge. As the numbers of *T. quinqueloba* were relatively low within the water column west of the Reykjanes Ridge, further sampling will be required to confirm this genotype distribution pattern and whether Type IIb is absent from the water column west of the Reykjanes Ridge.

4.3.3. Morphological variability of subarctic *T. quinqueloba*

Subarctic *T. quinqueloba* displayed a range of morphological variability (Plate 4-3). The specimens shown are from bulk plankton collections.

Plate 4-3

Plate 4-3 description: *Turborotalita quinqueloba* (Natland) from the subarctic Atlantic. The scale bar represents 100μm. Due to the mixed population of Type IIa and Type IIb genotypes found within the transect, they cannot be differentiated. **Fig. 1:** umbilical side, sinistral coiling, note small lip at central edge of aperture. **Fig. 2:** umbilical side, sinistral coiling, note lobate top chamber and flap that partially obscure the aperture. **Fig. 3:** umbilical side, dextral coiling, quadrate shape, lip at central edge of aperture. **Fig. 4:** umbilical side, sinistral coiling, umbilical flap obscures aperture. **Fig. 5:** umbilical side, sinistral coiling. **Fig. 6:** umbilical side, sinistral coiling, 5 chambers visible. **Fig. 7:** umbilical side, dextral coiling, rounded terminal aperture. **Fig. 8:** umbilical side, dextral coiling, very open chamber formation results in large aperture. **Fig. 9:** trochospiral side, sinistral coiling, note the ontogenic development of the chambers. Terminal chamber is smooth due to recent development. **Fig. 10:** juvenile specimen, umbilical side, dextral coiling, note apertural lip.

Specimens showed a range of morphological variation comparable to those outlined by Kroon *et al.* (1988). Basic morphological differences between specimens of *T. quinqueloba* include: the presence/absence of a lobate final chamber that obscures the aperture, large differences in test size, and some specimens also have a lip above the aperture.

Both sinistral and dextral coiling specimens were observed in the bulk collected specimens, and comparison of coiling direction from specimens east of the Reykjanes Ridge to specimens west of the Reykjanes Ridge showed that there was no significant difference between these areas when the error was considered (Fig. 4-12).

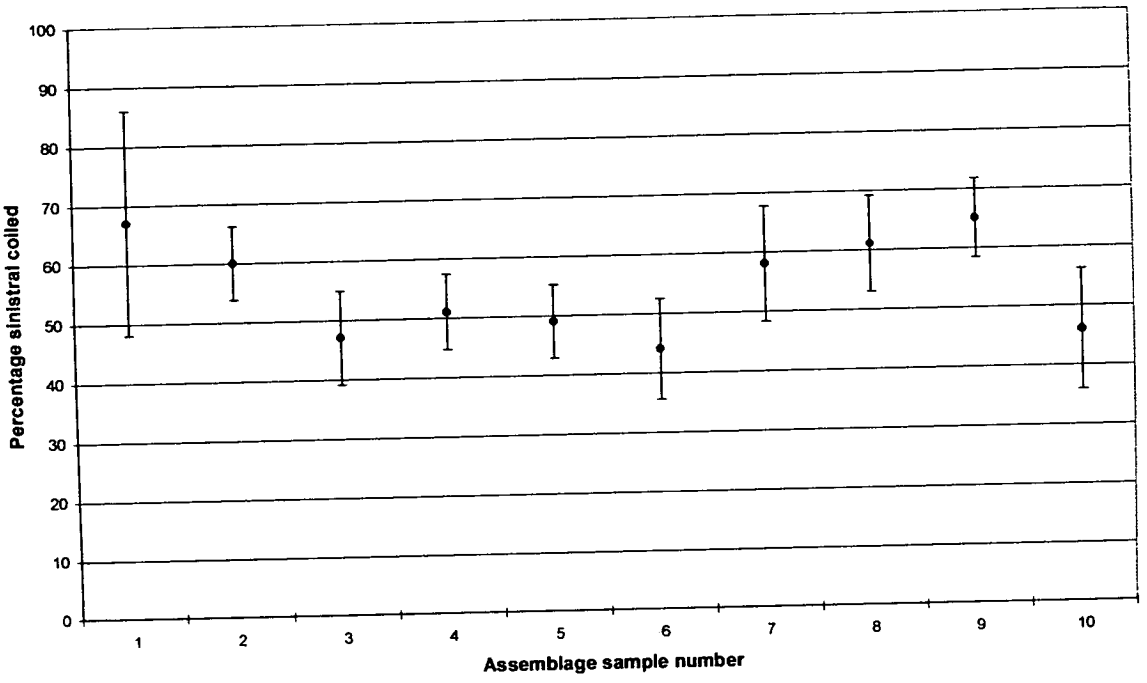


Fig. 4-12. Coiling differences within *T. quinqueloba* collected from ashed bulk plankton assemblages (see Table 2-3; based on a total of 403 specimens). Assemblages 1 and 2 are from east of Reykjanes Ridge, and assemblages 3 to 10 are from west of Reykjanes Ridge. The mean percentage of sinistrally coiled specimens for each assemblage is shown. In addition, the standard deviation from the mean for each assemblage is shown as an error bar. This takes into account the variation in specimen numbers obtained in each sample.

When the number of *T. quinqueloba* individuals sequenced (19) is compared to the number of individuals counted from the bulk plankton collections (403), there is a clear discrepancy and the low number sequenced should not be taken as representing the population density. This may be due to an artefact of the sampling procedure, as small foraminifera can be overlooked when quickly scanning samples for DNA analysis. Such selective sampling could account for the higher number of *G. bulloides* individuals picked as they are much larger and more noticeable. In addition, a number of specimens collected for DNA analysis were originally thought to be *T. quinqueloba* and turned out to be *N. pachyderma* (dextral). There is a morphotype of each morphospecies that is extremely similar in morphology, whilst viewing through a light

microscope, when the spines of *T. quinqueloba* have been stripped away during pump collection. However, they are easily distinguished with scanning electron microscopy on the basis of wall structure.

The coiling direction of the Type IIa genotypes sequenced was 75 % dextral (6 out of 8 specimens), and in contrast the coiling direction of the Type IIb genotypes was 90-100 % sinistral (9 out of 10 specimens; the coiling direction of one Type IIb specimen is not known). Comparison of the genotype/coiling direction ratios and the bulk plankton collection counts (Fig. 4-12), suggests that there is a mixed genotype population throughout the entire subarctic transect.

The cytoplasmic colouration was noted for the specimens sequenced. The Type IIa specimens were orange or yellow, and the Type IIb specimens were predominantly orange, although red and yellow were also observed.

4.4. The *Globigerina falconensis* cluster

A total of 50 *G. falconensis* specimens was obtained for phylogenetic analysis from the transitional-subtropical North Atlantic (see Figs. 2-1 and 2-3). Within the 505 bp molecular phylogeny (Fig. 3-2) the *Globigerina falconensis* cluster is supported in 100 % of the bootstrap replicates. The *G. falconensis* genotype from the North Atlantic is separated by an evolutionary distance of 0.8 % from the Coral Sea genotype within the 505 bp phylogeny (Fig. 3-2). Comparison of their SSU rDNA sequence (Fig. 4-13) shows that the genotypes have 4 base differences within the conserved regions used for phylogenetic analysis, with further differences within the variable regions of the fragment.

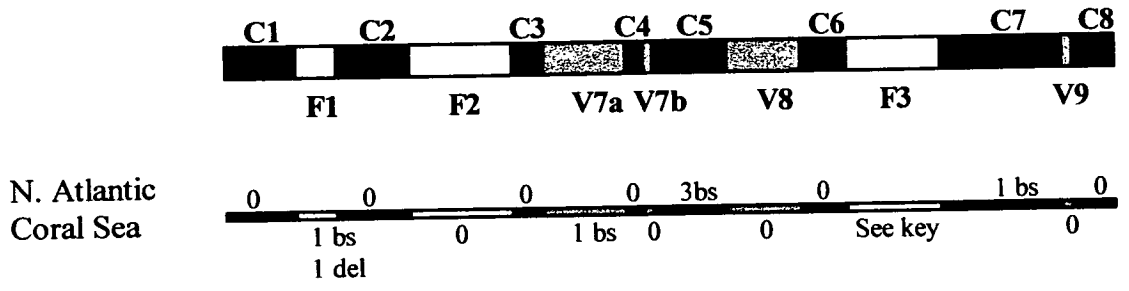


Fig. 4-13. Schematic representation of the ~1000 bp 3' terminal region of the SSU rRNA gene. Comparisons of the genetic differences are made between the *G. falconensis* genotypes from the North Atlantic and the Coral Sea. Indicated are the number of base differences between each of the genotypes, with the number of base substitutions and sequence length variations (base deletions/insertions) shown. **C1-C8** represent the highly conserved regions which were aligned relative to comparable regions present in all eukaryotes, **V7-V9** represent variable length expansion segments present in most eukaryotes and **F1-F3** represent three insertions which are unique to foraminifera. **bs** represents base substitutions and **del** represents sequence length variations. Within the region **F3** there are 99 aligned sites which have no base differences, and a region that is unalignable which is 60 and 48 bases in length for the North Atlantic and Coral Sea genotypes respectively.

In particular, the foraminiferal specific insertion F3 is very different between the two genotypes. There are 99 nucleotide sites which are completely aligned and have no base differences between the two genotypes. However there is a region which is completely unalignable with a sequence length variation of 12 bases between the two genotypes. Within the North Atlantic genotype the region has a length of 60 bases, but in the Coral Sea genotype it is only 48 bases. This indicates that the genotypes are far more divergent than the 505 bp phylogeny indicates. Although only a short evolutionary distance separates each of the genotypes, the relatively slow evolutionary rate within this lineage, reflected by the short branch length (see Chapter 3, section 3.4.1), suggests that this distance may actually represent a considerable period of time. Indeed, the level of sequence heterogeneity within the foraminiferal specific insertion

(F3) suggests that there may not have been gene flow between the two genotypes for a considerable period of time.

4.4.1. Sequence homogeneity of *G. falconensis* and distribution within the North Atlantic

A total of 50 SSU rDNA partial sequences was obtained from the North Atlantic (43 specimens from P247 and 7 specimens from M37/2). In contrast to both *G. bulloides* and *T. quinqueloba*, all *G. falconensis* specimens sequenced were identical throughout the entire SSU rDNA fragment. The distribution of the *G. falconensis* genotype is shown in Fig. 4-14. This morphospecies had one of the most extensive distributions of any species encountered during the Poseidon 247 collection, and was by far the most common *Globigerina* morphospecies obtained. It was identified within all three of the distinct water masses crossed during this collection, namely the North Atlantic Current (NAC), the North Atlantic Transitional Water (NATW) and the Azores Current (AC), over a sea surface temperature range of 12.8°C to 20.1°C.

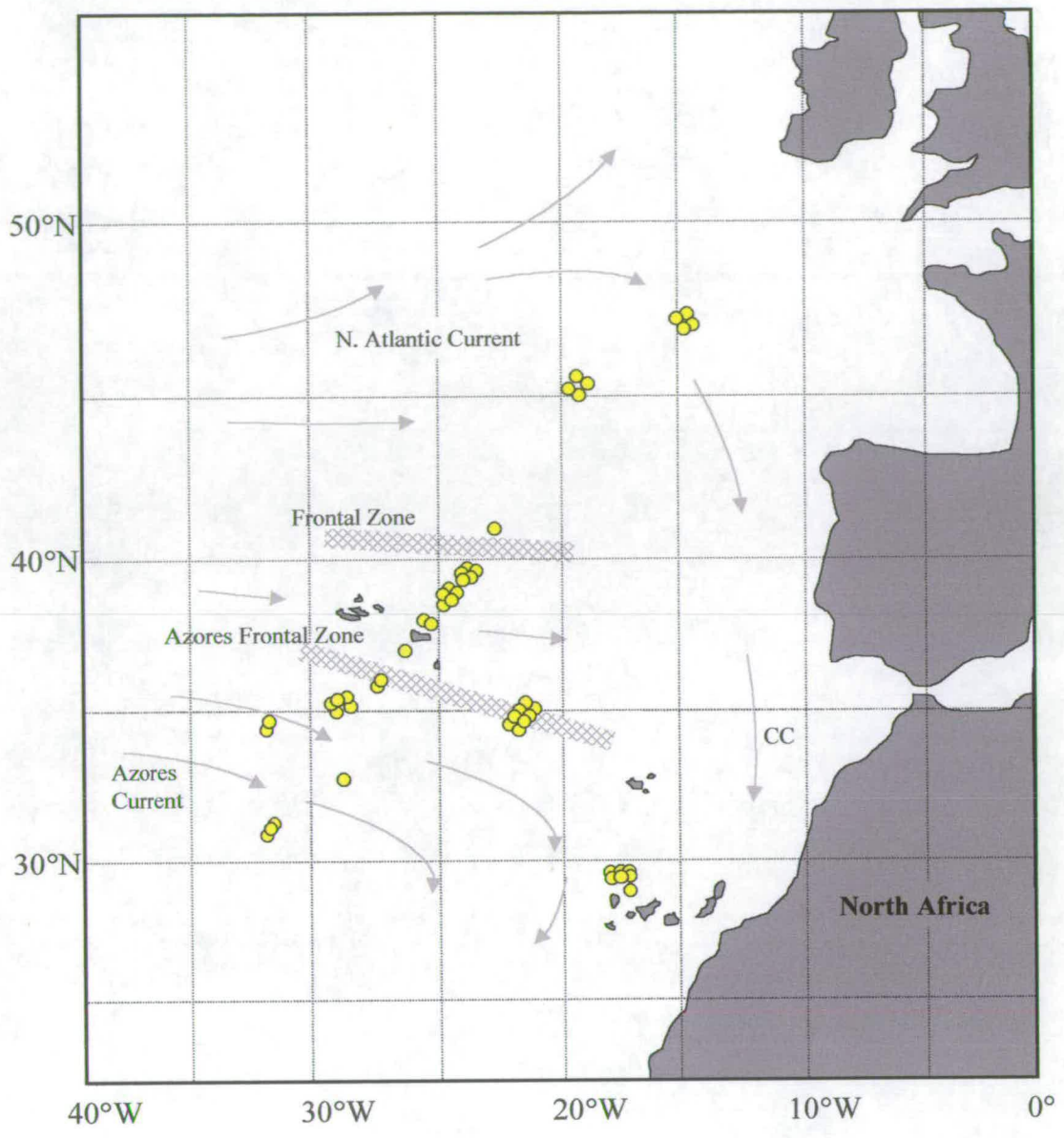


Fig. 4-14. The distribution of the *G. falconensis* genotype (●) within the North Atlantic, as found during collections M37/2 and P247. The major surface currents are indicated. **CC** represents the cool Canary Current. The approximate location of the main watermass frontal zones are shown, as determined by the shipboard thermo-salinometer.

4.4.2. Morphological variability of *G. falconensis*

The morphological variability of transitional-subtropical specimens of *G. falconensis* is shown in Plate 4-4.

Plate 4-4

Plate 4-4 description: *Globigerina falconensis* (Blow) from the subarctic Atlantic. The scale bar represents 100μm. **Fig. 1.** NAC specimen, umbilical side, sinistral coiling. **Figs. 2-3.** NAC specimens, umbilical side, dextral coiling. **Fig. 4-5.** NATW specimens, umbilical side, dextral coiling. **Fig. 6.** NATW specimen, umbilical side, sinistral coiling. **Fig. 7.** Azores frontal zone specimen, umbilical side, dextral coiling. **Fig. 8.** Azores frontal zone specimen, umbilical side, sinistral coiling. **Figs. 9-11.** AC specimens, umbilical side, sinistral coiling. **Fig. 12.** AC specimen, umbilical side, dextral coiling.

The specimens shown in Plate 4-4 are from bulk plankton samples, and have not all reached their mature size or form. There are a number of clear differences from the *G. bulloides* specimens shown in Plates 4-1 and 4-2. In general, the *G. falconensis* test is smaller than in the *G. bulloides* specimens. The test wall structure of *G. falconensis* is much more pustulose than found in *G. bulloides*. All the *G. falconensis* specimens have an apertural lip, and their aperture is much narrower than the open aperture of *G. bulloides*.

The numbers of *G. falconensis* specimens obtained from bulk plankton samples were too low to provide information regarding coiling direction. However, of the specimens sequenced, 68 % had sinistral coiling.

The cytoplasmic colouration was also noted and it was found that specimens were either yellow or orange. There was no correlation between geographic location and cytoplasmic colour.

4.5. Discussion

4.5.1. Distribution of *Globigerina/Turborotalita* genotypes within the North Atlantic

4.5.1.1. Genotype distribution in the subarctic region

Along the subpolar transect a distributional change was found, from a population composed predominantly of *G. bulloides* Type IIb genotypes (east of the Reykjanes Ridge) to a population composed entirely of *G. bulloides* Type IIa genotypes (west of Reykjanes Ridge) (Fig. 4-6). The transition from one population to the other corresponds approximately to the location of the Reykjanes Ridge, where a mixing zone is apparent between the two genotype populations. The mixing zone approximates to the boundary between the dominance of dextral to sinistral coiling of *N. pachyderma* observed in the sedimentary record by Ericson (1959). This probably represents a water mass transition between the influence of the relatively warm Irminger Current flowing west and the cold East Greenland Current flowing south (Fig. 2-2) and may also be influenced by the presence of the Reykjanes Ridge lying below. Two distinct genotypes were also identified within *T. quinqueloba* across the transect. However, these do not exhibit the same distribution pattern as *G. bulloides* (Fig. 4-11) as both genotypes co-exist to the east of the transect where the *G. bulloides* Type IIb population was found. Although only one genotype of *T. quinqueloba* (Type IIa) has so far been identified west of the Reykjanes Ridge within the *G. bulloides* Type IIa province, the low sample numbers do not provide sufficient evidence to conclude that a single genotype population of *T. quinqueloba* Type IIa is associated with the East Greenland Current. Indeed, comparison of *T. quinqueloba* genotype coiling direction ratios to the coiling direction of the bulk collected

specimens, suggests that the entire subarctic transect has a mixed genotype population.

4.5.1.2. Genotype distribution in the transitional-subtropical regions

In contrast to the subarctic region, *G. bulloides* represented a relatively small proportion of the planktic foraminiferal assemblage in the transitional-subtropical region during the January collection. Three distinct genotypes (Type Ia, Type IIa and Type IIb) were found co-existing within the North Atlantic Current (NAC) water mass. The Type Ia genotype was also found within the Canary Basin, and this genotype is almost identical to the genotype of de Vargas *et al.* (1997) from the Mediterranean Sea. The subpolar genotypes of *G. bulloides* (Types IIa and IIb) were found in the NAC (Fig. 4-7) and a single Type IIb specimen was even found within the subtropical waters of the Azores Current (AC) (35°N, Fig. 4-7).

The low numbers of *G. bulloides* found in the NAC during January is in contrast to its high abundance in this region during April and August (Ottens, 1991, 1992). Ottens (1992) also found that south of the Azores frontal zone *G. bulloides* tended to live at ~100m depth. As samples were collected by pumping from ~ 5m depth, it was thought that this could have accounted for only finding one specimen in the AC. However, samples taken between 700m and the surface using multi-nets showed that *G. bulloides* was absent from deeper in the water column in the AC (Schiebel, 1999). Also, quite surprisingly, no specimens of *T. quinqueloba* were found within the NAC or within the North Atlantic Transitional Water (NATW) during the January collection (P247). This is in contrast to the distributional studies of Ottens (1991, 1992) who found that *T. quinqueloba* was abundant in this region

during April and August. Both of these cases highlight the dramatic seasonal variability in morphospecies abundance and distribution patterns. In contrast to both *G. bulloides* and *T. quinqueloba*, the distribution of *G. falconensis* is represented by a single genotype. During the transitional-subtropical collection, *G. falconensis* was the most extensively found morphospecies, being found from the NAC through the NATW to the AC water (Fig. 4-14). Seasonality also has an effect on the distribution and abundance of *G. falconensis*. Schiebel (1999) noted that the abundance of *G. falconensis* within the AC system in January was much higher than in August. In addition, Ottens (1992) found *G. falconensis* as far north as 45°N in August, but in April only found *G. falconensis* south of the Azores frontal zone.

It is clear that each of the morphospecies exhibits a different genotype distribution pattern, and that within a morphospecies different genotypes may co-exist within the same watermass.

4.5.2. Relationship between genotype and the ocean environment

The relationship between genotype and water mass appears to be complex. Without further resampling, the possibility that the distribution patterns observed are a result of historical chance effects, or due to seasonality, cannot be ruled out. However, the genotypes representing *G. bulloides* and also *T. quinqueloba* do have different distributions. If this is a result of differing habitat preferences, based on a combination of temperature, salinity, nutrients or availability of food, then it would be of considerable significance to palaeoceanographers who use these morphospecies as proxies for palaeocean/climate reconstructions. In addition, the entire *G. falconensis* distribution is represented by a single genotype, suggesting that it is less

selective in its habitat preferences, or is tolerant to a range of oceanic conditions in the transitional-subtropical zone.

In the case of the *G. bulloides* genotypes, the subarctic data presented here suggests that temperature may influence their distribution. However, the Type IIa genotype mainly associated with the cooler water (as low as 2.5°C) found at the far east of the subarctic transect (Fig. 4-6), has also been identified within the warmer water of the NAC (14.2°C) as far south as 43°N (Fig. 4-7). Further the Type IIb genotype from the subarctic region was also found in the subtropical water of the AC (19.3°C) (Fig. 4-7). Both Type IIa and Type IIb are apparently tolerant to a wide temperature range, suggesting that other factors are also important in the distribution of *G. bulloides* genotypes. A lack of correlation between temperature and genotype distribution is also observed in *T. quinqueloba*, where two genotypes are found living sympatrically within the same water mass (Fig. 4-11). Further, a single *G. falconensis* genotype was found across 3 distinct water masses, over a temperature range of <13°C to ~20°C. This temperature range is not as great as observed in the distribution of *G. bulloides* Types IIa and IIb. The distribution of *G. falconensis* appears to be limited by intolerance to cold water, since its distribution in the N. Atlantic does not extend into the subarctic waters (Ottens, 1992).

Salinity is likely to play an important role in planktic foraminiferal distribution in the subarctic region. Although not measured, the salinity along the East Greenland margin is lower than further offshore (Fig. 4-15; Levitus *et al.*, 1994) due to the East Greenland Current having a low salinity (Carstens and Wefer, 1992; Carstens *et al.*, 1997) because of the input of glacial meltwater along the east

Greenland margin. It is possible that this influences the distribution of foraminiferal genotypes in this region.

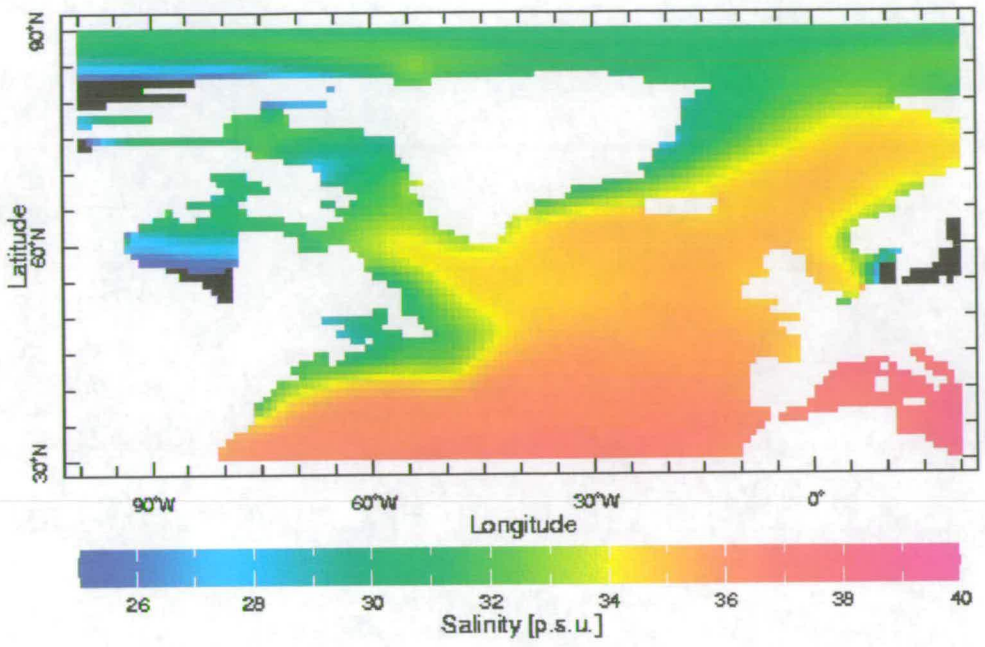


Fig. 4-15. Salinity of the North Atlantic Ocean during August (Levitus *et al.*, 1994).

Productivity levels within the subpolar region are at their height during the collection period (August and September), and at a relatively low level during the transitional-subtropical collection (January). The productivity of the North Atlantic is illustrated in Fig. 4-16. However, phytoplankton distribution is affected by mesoscale eddies (Washburn *et al.*, 1998), which creates phytoplankton patchiness (Fig. 4-16) as recorded by Gower *et al.* (1980) in the region south of Iceland. Individual foraminiferal morphospecies are known to thrive at different times within the cycle of a phytoplankton bloom (Kroon and Ganssen, 1989) suggesting that they may have different feeding preferences. Patchy phytoplankton distribution and feeding preferences may account for the observed genotype distribution patterns. The low number of *G. bulloides* found in the NAC during January, when the primary

productivity was relatively low, is perhaps due to *G. bulloides* preferring higher production conditions.

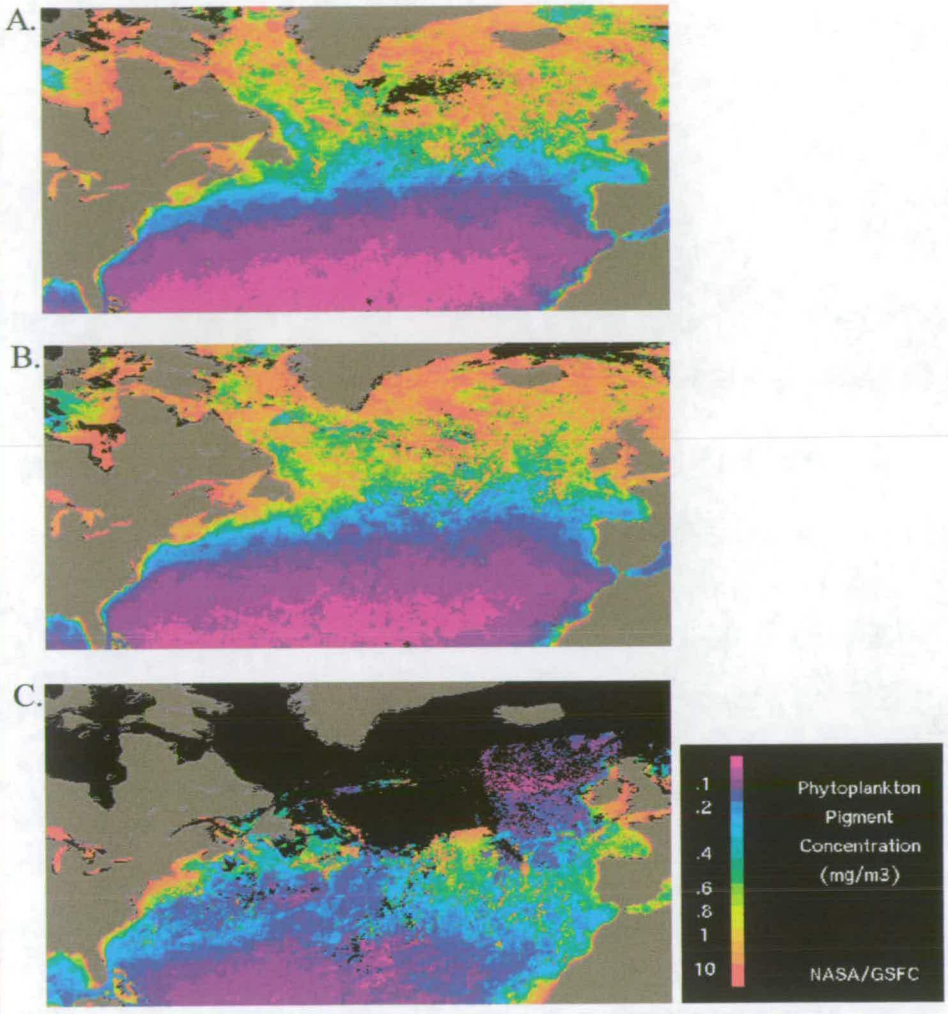


Fig. 4-16. Satellite imagery of primary production within the North Atlantic during August (A), September (B) and January (C). The scale indicates the relative phytoplankton concentration. From NASA/GSFC SeaWifs Ocean Colour website (http://daac.gsfc.nasa.gov/CAMPAIGN_DOCS/OCDST/ocdst_north_atlantic_productivty.html).

It was predicted that cytoplasmic colour could possibly provide a potential indicator of feeding preference. However, the lack of correlation between genotype distribution and cytoplasmic colour found during this study indicates that variations

in colour are probably due to a combination of phytoplankton patchiness, opportunistic feeding behaviour or the stage of digestion of food vacuoles, and not related to foraminiferal genotype. Cytoplasmic colour therefore does not provide a proxy for genotype. If the distribution of genotypes is due to productivity patchiness, the precise relationship between genotype and productivity will only be resolved with *in-situ* nutrient and chlorophyll measurements at the time of collection.

As the foraminiferal population was not continually sampled, it is possible that further genotypes remain to be discovered. The observed distribution patterns may also vary temporally and spatially due to seasonal changes in oceanographic parameters, as is apparent for *T. quinqueloba* within the transitional-subtropical region. However, low sampling numbers of *G. bulloides* and *T. quinqueloba* must indicate that the genotypes discovered must be present in reasonable numbers. Further, the findings conclusively demonstrate that genotype distribution patterns are species specific, and clearly not only related to temperature.

4.5.3. Morphology/genotype relationships

Having established that planktic foraminiferal morphospecies often represent complexes of distinct genotypes, the next question to answer is whether the genotypes have distinctive morphological characteristics. As the numbers of mature specimens collected were very low, morphometric analysis was not used since it would not have been statistically significant. Plate 4-1 shows the morphotypes of subarctic *G. bulloides* that represent the Type IIa and Type IIb populations respectively. The density of sampling is sufficient to assume that morphotypes 1-5 in Plate 4-1 are representative of *G. bulloides* Type IIa, with morphotypes 7-12

representative of *G. bulloides* Type IIb. I was unable to identify any clear morphological differences characteristic of the two subarctic *G. bulloides*. It is thus not possible to identify these genotypes in the sediment although further advanced morphometric analysis (e.g. image analysis), porosity calculations from mature specimens obtained from sediment traps, and chemical and isotope analysis, may enable their discrimination.

The morphology of the transitional-subtropical *G. bulloides* specimens (Plate 4-2) is comparable to the subarctic specimens. There were three distinct genotypes in this region, but since low numbers were collected from bulk plankton samples it was impossible to try and discriminate them. One *G. bulloides* Type Ib specimen sequenced had an apertural lip. This has the potential to be confused with specimens of *G. falconensis*. Additional investigation will be required to determine whether this is a common morphological characteristic of this genotype.

Plate 4-3 (1-10) illustrates the wide range of *T. quinqueloba* morphotypes found throughout the subarctic transect. Sampling density of specimens for DNA analysis was not sufficient to determine whether Type IIb was also present west of the Reykjanes Ridge. In contrast to the apparent lack of morphological variation within *G. bulloides*, *T. quinqueloba* displays an extensive variety of morphological form. Most notable is the similarity between morphotypes of *T. quinqueloba* (Plate 4-3) and *N. pachyderma* (Plate 7-1) mentioned by Bé and Tolderlund (1971), and which could be easily confused in sediment samples without close scrutiny. The Type IIa genotypes have both sinistral and dextral coiling specimens. The Type IIb genotypes are all sinistral coiling (except perhaps for 1 specimen, for which I do not know the coiling direction), although this may be chance and a function of the low

numbers. This warrants further investigation, and by combining morphometric analysis and porosity calculations from mature specimens obtained from sediment traps together with a more extensive genetic investigation, it may be possible to resolve the genotype/morphotype relationship.

The morphological variation observed in *G. falconensis* is shown in Plate 4-4. As only a single genotype was found in the transitional-subtropical region, any morphological variation must reflect phenotypic plasticity. The number of specimens obtained from bulk plankton samples were relatively low, and they were not all mature specimens. This has prevented investigating whether the phenotypic plasticity is a result of natural morphological variation within a population, or whether the variations are environmentally induced.

Since genotypes of *G. bulloides* and *T. quinqueloba* often co-exist, it was impossible to investigate whether morphological variation in these morphospecies maps to genotype.

4.5.4. Cryptic speciation within *G. bulloides* and *T. quinqueloba*

Cryptic species of planktic foraminifera are by definition difficult to distinguish by traditional paleontological methods (Huber *et al.*, 1997). However, using a combination of morphological, biological, isotopic and genetic data, Huber *et al.* (1997) concluded that *Globigerinella siphonifera* Types I and II were cryptic species. The existence of three cryptic species, which have considerable genetic distance, has also been proposed for *O. universa* (Darling *et al.*, 1999; de Vargas *et al.*, 1999).

Molecular phylogenetic analyses of the foraminifera have revealed that planktic foraminiferal morphospecies are commonly characterised by complexes of SSU rDNA genotypes (Darling *et al.*, 1999; de Vargas *et al.*, 1999). Similarly, this has now been observed in *G. bulloides* and *T. quinqueloba*. The division within both *G. bulloides* and *T. quinqueloba* of two main clusters (Types I and II) may be indicative of cryptic speciation events, similar to those described above. Initial indications suggested that the *G. bulloides* and *T. quinqueloba* genotypes are adapted to warm (Type I) and cool (Type II) water. However, the genotype distribution data from the transitional-subtropical collection shows this to be an over-simplified explanation for *G. bulloides*. The "cool water" specimens appear to be tolerant to a wide range of ocean conditions, but it may be possible that "warmer water" types may not be adapted to live in cold water masses. This will only be resolved with further investigation.

Within the Type II clusters, a relatively small genetic distance separates *G. bulloides* Type IIa from IIb and *T. quinqueloba* Type IIa from IIb within the conservative 505bp phylogeny (Fig. 3-2). However, within the variable regions of the SSU rDNA fragment amplified there is considerable sequence variation (Fig. 4-5 and Fig. 4-10). It is also possible that some of the variants observed within the cool water genotypes may represent cryptic species, since it is already recognised that sibling species may show very little genetic differentiation (Knowlton, 1993).

Further work will be required to determine which, if any, of the genotypes within these complexes represent cryptic species. Although at present there is no definitive evidence, cryptic speciation or genotype adaptation to a specific habitat cannot be ruled out. The larger evolutionary distance that separates Type I and II is

perhaps more supportive of cryptic speciation than the small evolutionary distance within the Type II clusters.

This has important consequences for palaeoceanographic reconstructions that utilise *G. bulloides*. The distribution patterns are obviously complex, and it is possible that the genotypes are adapted to a specific habitat preference. The problem is highlighted by the fact that three distinct *G. bulloides* genotypes co-exist within the NAC water. Although cryptic speciation and genotype habitat adaptation cannot yet be proven, neither can it be discounted. If these genotypes are adapted to some different aspect of the ocean environment, then grouping them together for palaeoceanographic investigations could possibly introduce a significant amount of noise, or error, into these studies. By linking genotype/morphotype relationships, examining genetic variation within other genes, and determining whether genotypes are actually adapted to specific habitats, this has the potential to provide new high resolution proxies from the marine sediment record.

4.5.5. Is *Globigerina falconensis* of palaeoceanographic importance ?

In contrast to the genetic variation observed within *G. bulloides* and *T. quinqueloba*, *G. falconensis* showed no genotypic variation within the North Atlantic specimens sequenced. During the transitional-subtropical collection, *G. falconensis* was the most extensively found morphospecies, being found from the NAC through NATW to the AC water. It could be argued that *G. falconensis* would be more advantageous to use than *G. bulloides* for palaeoceanographic reconstructions that

utilise isotopes. In discussing the use of *G. falconensis*, the following points must be considered:

1. *G. falconensis* within the transitional-subtropical region of the N. Atlantic is represented by a single genotype, hence there is no indication of cryptic speciation or adaptation.
2. *G. falconensis* has an extensive distributional range within the transitional-subtropical zone, and has a high relative abundance.
3. *G. falconensis* is thought to live nearer the surface than *G. bulloides* (Malmgren and Kennett, 1977) and would therefore reflect sea surface temperatures (SSTs) more accurately.
4. *G. bulloides* is known to secrete its test out of equilibrium with respect to carbon isotopes (Deuser and Ross, 1989; Sautter and Thunell, 1991b; Kroon and Darling, 1995; Spero and Lea, 1996), which can hinder interpretation of stable isotope data.
5. The distribution of *G. falconensis* does not reach the high latitudes, although it possibly has done so during extreme warm climates.
6. *G. falconensis* is difficult to distinguish from *G. bulloides*. However this is no more of a problem than *vice versa*.
7. *G. falconensis* is symbiotic which may cause carbon isotope fractionation due to symbiont photosynthesis as found in *G. siphonifera* (Bijma *et al.*, 1998). Other chemical proxies such as Cd/Ca ratios may be suitable.

It is difficult to say whether *G falconensis* would be an important addition to palaeoceanographic isotope proxies without further work, however it has some significant advantages over *G. bulloides*. It would therefore be very interesting to see

how isotope-based estimated SSTs for the transitional-subtropical region of the North Atlantic would compare between *G. bulloides* and *G. falconensis*.

4.5.6. Gene flow

One of the most significant findings is that two genotypes of *G. bulloides* and one *T. quinqueloba* genotype (Type IIa) are identical between the subarctic and the subantarctic. In addition, a further subarctic *T. quinqueloba* genotype (Type IIb) was found to be very closely related to another subantarctic *T. quinqueloba* genotype (Type IIc). These morphospecies display a predominantly anti-tropical distribution, but the data show that the genotypes have recently mixed across the tropics, between the cool high latitude oceans. This shall be examined in detail in Chapter 8.

The *G. bulloides* Type Ib genotype from the North Atlantic is almost identical to the Mediterranean genotype of de Vargas *et al.* (1997). However, there are possible errors within the Mediterranean sequence and the genotypes are most likely identical. This would indicate gene flow between the North Atlantic and the Mediterranean. Since the surface current systems flow from the North Atlantic to the Mediterranean, rather than vice versa, this would suggest that gene flow must be predominantly in that direction. However, gene flow could only occur in the other direction if planktic foraminifers would survive transiting in saline intermediate water that flows from the Mediterranean into the North Atlantic. This may be possible since foraminifers live in the eastern Mediterranean where the salinities are much higher.

An evolutionary distance of 4.7 % separates the *G. bulloides* Type Ia (Coral Sea) and Type Ib (North Atlantic) genotypes, suggesting that the two populations

have not mixed for a considerable period of time. Similarly, although a relatively small evolutionary distance separates the *G. falconensis* genotypes from the North Atlantic and the Coral Sea, the slow evolution rate within this lineage reflected in its relatively short branch suggests that this may actually represent a considerable period of time (millions of years ?). Gene flow patterns between different areas of the globe are discussed in greater detail in Chapter 8, where other taxa are also considered.

Chapter 5: *Globigerinella* sp. Index

5.1. Introduction to species within this lineage

In this chapter the *Globigerinella* clade is examined, which comprises the morphospecies *Globigerinella siphonifera* d'Orbigny and *Globigerinella calida* (Parker). Collectively, these morphospecies inhabit the transitional and subtropical/tropical water masses, and represent an important part of the planktic foraminiferal assemblage. Although an important component of transfer functions for the calculation of palaeo-SSTs, there has been limited use of these morphospecies by paleoceanographers. However, comprehensive investigations have recently been carried out on morphotypes of *G. siphonifera* providing a better understanding of their biology (Huber *et al.*, 1997; Bijma *et al.*, 1998). This chapter starts by outlining the present day distribution pattern of the *Globigerinella* morphospecies, and reviews previous work. The phylogenetic relationship between, and within, each of the morphospecies is then examined, and the distribution of *Globigerinella* genotypes within the North Atlantic is described. This is followed by an examination of morphological variability found within the *Globigerinella* morphospecies. The chapter is completed with a discussion of the results.

5.1.1. *Globigerinella siphonifera* (d'Orbigny)

At present, *G. siphonifera* has an extensive distribution within the oceans being found from temperate to tropical water masses across a temperature range of between 10-30°C (Tolderlund and Bé, 1971). *Globigerinella siphonifera* prefers to live in boundary currents, areas of upwelling and near continental margins (Bé and Tolderlund, 1971). Within the North Atlantic, Ottens (1992) found that the *G.*

siphonifera abundance maxima occurred between 55-50°N in August, and between 35-30°N in April.

Based on observations by Faber *et al.* (1988, 1989), extensive research into *G. siphonifera* has provided considerable biological evidence for the support of species level division into two distinct types. The two cryptic species are distinguished on the basis of their genetic, morphological, biological, and chemical differences (Darling *et al.*, 1997; Huber *et al.*, 1997; Bijma *et al.*, 1998): (1) Molecular distances – the average evolutionary distance between Types I and II is greater than 6 % in a molecular phylogeny based on 604 bp (Darling *et al.*, 1997); (2) Morphological differences – biometric differences and porosity variations. Adult specimens of Type I have an average porosity of 10-30 %, where as adult specimens of Type II have an average porosity of 4-10 % (Huber *et al.*, 1997); (3) Biological differences – Types I and II differ in their rhizopodial network and spine distribution (Faber *et al.*, 1988; Huber *et al.*, 1997); (4) Isotopic differences – the isotope values indicate that Types I and II precipitate their calcite at a temperature difference of between 1.5-1.7°C, with Type I precipitating its calcite at the higher temperature (Bijma *et al.*, 1998); (5) Endosymbionts – Type I has facultative symbionts, which are low in number, whereas Type II has obligatory symbionts, which are high in number (Faber *et al.*, 1988, Bijma *et al.*, 1998).

The distribution of Types I and II is poorly understood. It is unknown whether *G. siphonifera* Type I exists in the Pacific Ocean, nor how Types I and II are distributed within the North Atlantic. Further, it is not clear when the divergence between Types I and II occurred (Huber *et al.*, 1997). In addition, by examining the distribution of *G. siphonifera* genotypes in the North Atlantic, further insight may be

gained into the preferred habitats of both *G. siphonifera* Type I and Type II, which may enable examination of the processes that lead to their speciation.

5.1.2. *Globigerinella calida* (Parker)

Globigerinella calida is thought to be mainly distributed within tropical to subtropical water masses (Kennett and Srinivasan, 1983). However, within the North Atlantic Funnell and Swallow (1997) recorded the presence of *G. calida* in core tops from the transitional waters of the North Atlantic Current at ~50°N. In the South Pacific, *G. calida* is found in the warm waters north of 35°S near New Zealand, and north of 30°S in the east Pacific (Parker, 1962).

Juvenile specimens of *G. calida* and *G. siphonifera* are very difficult to distinguish (Hemleben *et al.*, 1989), although mature specimens are more obviously morphologically different since *G. siphonifera* has planispiral coiling, whereas *G. calida* is more trochospiral. Initial descriptions of *G. calida,* by Bé and Tolderlund (1971), suggested that *Globigerinella calida* is an intermediate form between *G. bulloides* and *G. siphonifera*. For practical purposes, they lumped "calida-like" forms with *G. siphonifera.* However, Kennett and Srinivasan (1983) determined that *G. calida* most likely evolved from *G. siphonifera* and that it had no association to *G. bulloides.*

Globigerinella calida has not been used specifically in paleoceanographic studies, probably due in part to the difficulty in discriminating juvenile specimens of *G. calida* and *G. siphonifera* in the 125-250 μm size fraction, and also its relatively recent appearance in the fossil record (~ 4 Ma, Kennett and Srinivasan, 1983).

Indeed, North Atlantic sea-surface temperature estimates based on transfer functions by Funnell and Swallow (1997) lumped both morphospecies together.

Recently, de Vargas *et al.* (1997, 1998) attempted to use *G. calida* to calibrate a molecular clock within the planktic spinose region of the foraminiferal molecular phylogeny. Unfortunately, further phylogenetic analyses (Kate Darling, pers. comm.) showed that they misidentified "*G. calida*", and the specimen they used for analysis was actually *G. siphonifera* Type II. The specimen is identical to the *G. siphonifera* Type IIa genotype obtained from the North Atlantic during this study. This meant that their "clock" calculations were based on an incorrect calibration datum. As discussed in Chapter 3, there is significant discrepancy regarding the timing of the *G. calida* divergence from the *Globigerinella* lineage, between the interpretation of the fossil record (Kennett and Srinivasan, 1983) and the molecular phylogeny (Fig. 3-2). The molecular phylogeny suggested that the first appearance of *G. calida* was 10-11 Ma (Chapter 3, section 3.4.2), compared with the interpretation of the fossil record which suggested that the first appearance was 4 Ma (Kennett and Srinivasan, 1983). Until a definitive divergence datum can be ascertained, *G. calida* cannot not be used to calibrate the *Globigerinella* lineage.

5.2. Phylogenetic relationships within the *Globigerinella* genotype cluster

A total of 42 *G. siphonifera* specimens was collected for phylogenetic analysis from the transitional-subtropical region of the North Atlantic during Poseidon 247 and Meteor 38/2 cruises (see Chapter 2, Figs. 1 and 3). A total of 17 specimens of *G. calida* was collected for phylogenetic analysis from the transitional-subtropical region of the North Atlantic during the Poseidon 247 cruise (see Fig. 2-

3). The phylogeny presented in Fig. 5-1 represents a sub-tree of the more extensive 505 bp foraminiferal molecular phylogeny (Fig. 3-2). The tree illustrates the phylogenetic position of the North Atlantic *G. siphonifera* Type I, *G. siphonifera* Type II, and *G. calida* genotypes obtained during this study.

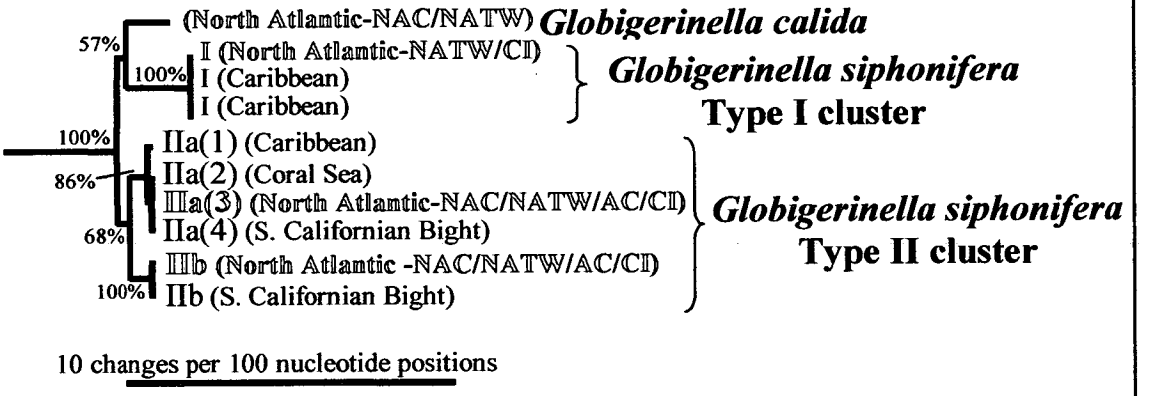


Fig. 5-1. Sub-tree of neighbour-joining phylogeny based on 505 unambiguously aligned sites, showing the evolutionary relationships between the *Globigerinella* morphospecies. The *G. siphonifera* Type I and II clusters are indicated together with the phylogenetic position of *G. calida*. Support has been assigned to the phylogeny by using 1000 bootstrap replicates. The region of the North Atlantic from which the genotypes were obtained is shown. **NAC** the North Atlantic Current, **NATW** the North Atlantic Transitional Water, **AC** the Azores Current and **CI** the Canary Islands.

The *Globigerinella* branch divides into two clusters, with each having a bootstrap support of 57 % and 68 % respectively. The first cluster includes *G. siphonifera* Type I (NATW and Canary Islands) and *G. calida* (NAC and NATW). Their association within the phylogeny has not been resolved fully as the branch only has support from 57 % of the bootstrap replicates. The second cluster consists of *G. siphonifera* Type II genotypes from the North Atlantic (NAC, NATW, AC and Canary Islands), the Southern Californian Bight, the Caribbean, and the Coral Sea.

The clustering of the Type IIa genotypes within a single lineage is supported in 68 % of the bootstrap replicates, a level of support which is close to the accepted significance level of 70 % (Hillis and Bull, 1993). The cluster topology (Fig. 5-1) indicates that *G. siphonifera* Type I, Type II, and *G. calida* all diverged from the *Globigerinella* lineage over a relatively short period of time.

To attempt to resolve the relationships within the *Globigerinella* cluster, a phylogeny of only *Globigerinella* genotypes was constructed. This neighbour-joining phylogeny was based on 767 unambiguously aligned sites and is presented in Fig. 5-2.

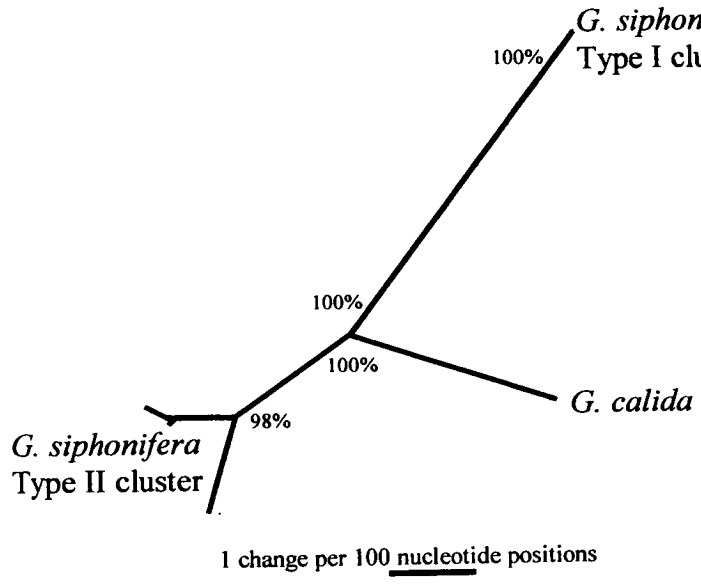


Fig. 5-2. Neighbour-joining phylogeny based on 767 unambiguously aligned sites, showing the evolutionary relationships between the *Globigerinella* morphospecies. Support has been assigned to the phylogeny using bootstrap 1000 replicates.

The 767 bp neighbour-joining phylogeny (Fig. 5-2) illustrates that *G. siphonifera* Types I and II genotype clusters are supported in 100 % and 98 % of the bootstrap replicates respectively, and are separated by a mean evolutionary distance

of 7.0 %. The *G. siphonifera* Type I cluster is separated from the *G. calida* genotype by a mean evolutionary distance of 6.9 %, with their association supported in 100 % of the bootstrap replicates. The *G. siphonifera* Type II cluster is separated from the *G. calida* genotype by a mean evolutionary distance of 5.1 %, with their association supported in 100 % of the bootstrap replicates.

5.2.1. Sequence length variations between *Globigerinella* genotypes

Within the *Globigerinella* cluster there is considerable sequence length variation within the region of the SSU rRNA gene amplified. The *G. siphonifera* Type I genotype has a sequence length of 1032 bp, the *G. siphonifera* Type II genotypes have sequence lengths of between 1003-1015 bp and the *G. calida* genotype has a sequence length of 1020 bp.

5.2.2. *Globigerinella siphonifera* Type I

The *G. siphonifera* Type I cluster has support from 100 % of the bootstrap replicates (Fig. 5-2). The Type I cluster consists of genotypes from the North Atlantic (this study) and the Caribbean (Darling *et al.*, 1997; de Vargas *et al.*, 1997). The North Atlantic genotype was found to be genetically identical to both of the Caribbean genotypes not only within the conserved regions used for phylogenetic analysis, but also throughout the variable regions of the SSU rDNA fragment.

5.2.3. *Globigerinella siphonifera* Type II (d'Orbigny)

The *G. siphonifera* Type II cluster has support from 98 % of the bootstrap replicates (Fig. 5-2). To illustrate the phylogenetic relationships between the Type II

genotypes more clearly, a neighbour-joining phylogeny was constructed with the Type II genotypes and using the *G. siphonifera* Type I (North Atlantic) genotype as an outgroup. This phylogeny is presented in Fig. 5-3.

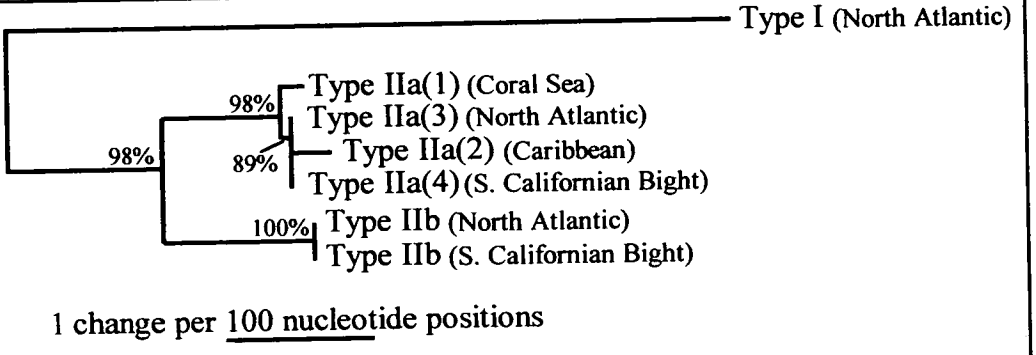


Fig. 5-3. Neighbour-joining phylogeny based on 767 unambiguously aligned sites, illustrating the phylogenetic relationships within the *G. siphonifera* Type II cluster. The *G. siphonifera* Type I (North Atlantic) has been used as an outgroup. Support has been assigned to the phylogeny using 1000 bootstrap replicates.

The *G. siphonifera* Type II cluster further divides into two distinct clusters (Type IIa and Type IIb). The Type IIa and Type IIb clusters have bootstrap support of 98 % and 100 % respectively. A mean evolutionary distance of 2.1 % separates the Type IIa and Type IIb clusters.

Within the North Atlantic two distinct *G. siphonifera* Type II genotypes have been identified which fall into the Type IIa and IIb clusters respectively. A total of 28 specimens of Type IIa and 11 specimens of Type IIb were obtained from the North Atlantic collections. An evolutionary distance of 2.0 % separates the two genotypes from the North Atlantic.

The Type IIa cluster is formed from genotypes from the N. Atlantic, Southern Californian Bight, Coral Sea and the Caribbean. Within the Type IIa cluster the

genotypes from the North Atlantic, the Caribbean and the Southern Californian Bight cluster separately from the Coral Sea genotype with high bootstrap support (89%). The evolutionary distances between the genotypes within the Type IIa cluster are low, representing relatively few base differences within the regions used for phylogenetic analysis.

The Type IIb cluster consists of genotypes from the N. Atlantic and the Southern Californian Bight. The two genotypes show complete sequence homogeneity, with not a single base difference in the ~1000 bp fragment.

As relatively low evolutionary distances separate the *G. siphonifera* Type II genotypes, a comparison of the entire ~1000 bp SSU rDNA fragment between the Type II genotypes is shown in Fig. 5-4. The North Atlantic genotypes are compared against each other, and against the other genotypes within their respective clusters. The number of base substitutions and base deletion/insertions are shown. An explanation of these terms can be found in Chapter 4 (section 4.3.2., Fig. 4-4).

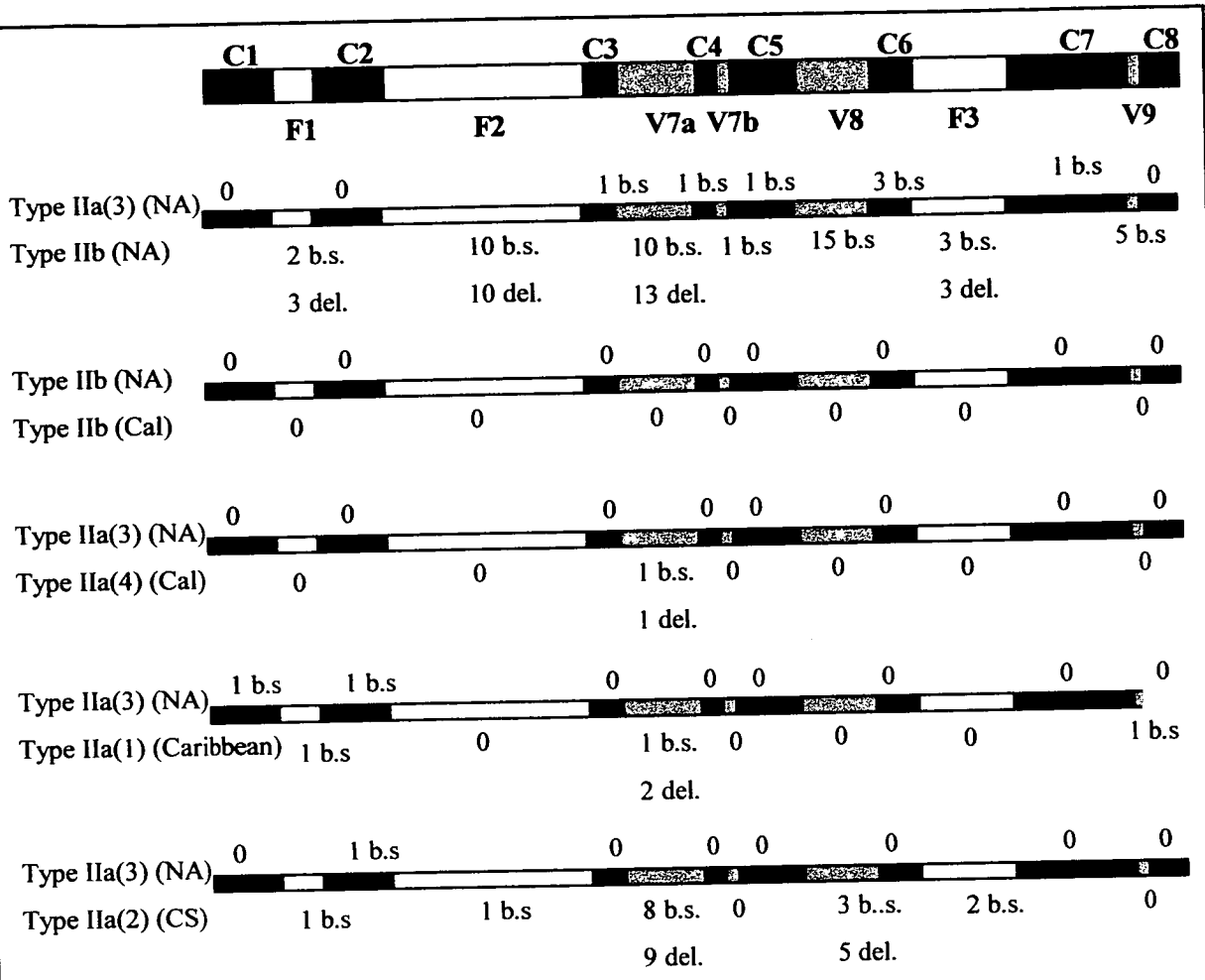


Fig. 5-4. Schematic representation of the ~1000 bp 3' terminal region of the SSU rRNA gene. Comparisons of the genetic differences are made between the genotypes from the North Atlantic, Southern Californian Bight, Caribbean, and Coral Sea. The number of base differences between each of the genotypes are indicated, with the number of base substitutions and sequence length variations (base deletions/insertions) shown. **C1-C8** represent the highly conserved regions which were aligned relative to comparable regions present in all eukaryotes, **V7-V9** represent variable length expansion segments present in most eukaryotes and **F1-F3** represent three insertions which are unique to foraminifera. **bs** represents base substitutions and **del** represents sequence length variations due to insertions/deletions. **NA** denotes the North Atlantic, **Cal** denotes the S. Californian Bight, and **CS** denotes the Coral Sea.

It is apparent that the North Atlantic Type IIa(3) genotype is very closely related to the Southern Californian Bight Type IIa(4) genotype, with only 1 base substitution and 1 base deletion across the entire ~1000 bp SSU rDNA fragment

(region V7a, Fig. 5-4). A slightly greater level of sequence divergence is observed between the N. Atlantic Type IIa(3) and Caribbean Type IIa(1) genotypes (Fig. 5-4). Within the conservative regions C1 and C2, the genotypes differ by two base substitutions, with a further 3 base substitutions and 2 base insertion/deletion differences occurring within the variable regions. A much higher level of sequence divergence is observed between the N. Atlantic Type IIa(3) and Coral Sea Type IIa(2) genotypes, with a single base substitution within the conserved region C2, in addition to 15 base substitution and 14 base insertion/deletion differences within the variable regions (Fig. 5-4).

The evolutionary distance of 1.98 % separating the N. Atlantic *G. siphonifera* Type IIa(3)and Type IIb genotypes represents 16 substitutional base differences within the 767 bp used for phylogenetic analysis. Across the entire ~1000 bp SSU rDNA fragment, the North Atlantic Type IIa(3) and Type IIb genotypes differ by a total of 53 base substitutions, and 29 base deletions/insertions (Fig. 5-4). The detailed genotype analysis highlights the large sequence divergence that exists between the N. Atlantic Type IIa(3) and Type IIb genotypes, which is not as evident when only the molecular phylogeny is considered. It is clear that the often co-existing (see distribution section 5.3.2.) North Atlantic Type IIa(3) and IIb genotypes have a much closer genetic relationship to the other genotypes within their respective clusters, than to each other.

5.2.4. *Globigerinella calida*

Partial SSU rDNA sequences were obtained from a total of 17 specimens of *G. calida*. All were genetically identical, except for two specimens from the Azores

Current region (see *G. calida* distribution patterns, Fig. 5-6) which had 3 base substitutions and 1 base insertion/deletion within the variable regions of the SSU rDNA fragment (Fig. 5-5). Only a 500 bp fragment was successfully amplified from the two variant specimens and further base differences may exist in the other section of the SSU rRNA gene fragment.

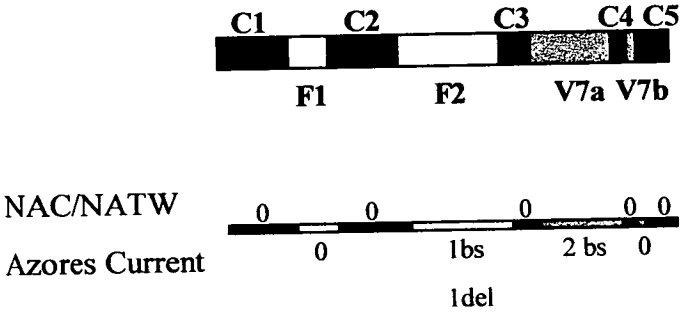


Fig. 5-5. Schematic representation of an ~500 bp region of the SSU rRNA gene. Comparisons of the genetic differences are made between the *G. calida* genotypes from the North Atlantic Current (NAC) / North Atlantic Transitional Water (NATW) and the Azores Current regions of the North Atlantic. The number of base differences between each of the genotypes are indicated, with the number of base substitutions and sequence length variations (base deletions/insertions) shown. **C1-C5** represent the highly conserved regions which were aligned relative to comparable regions present in all eukaryotes, **V7a-V7b** represent variable length expansion segments present in most eukaryotes and **F1-F2** represent three insertions which are unique to foraminifera. **bs** represents base substitutions and **del** represents sequence length variations.

5.3. Distribution of *Globigerinella* genotypes within the transitional-subtropical assemblage zones of the North Atlantic

The distribution of *Globigerinella* genotypes, sampled during the Poseidon 247 and Meteor 38/2 cruises, is described in the following section. The distribution of each morphospecies is dealt with separately, and is illustrated in Fig. 5-6 and Fig. 5-7.

5.3.1. Distribution of *Globigerinella siphonifera* Type I

Only 3 specimens of *G. siphonifera* Type I were obtained during the transitional-subtropical collections. One specimen was obtained during Poseidon 247 from the North Atlantic Transitional water (NATW), between the Azores island of San Miguel and the Azores frontal zone (Fig. 5-6). The water temperature at this locality was 17.4°C. Two further specimens were obtained during Meteor 38/2 from the ESTOC area (Fig. 5-6), north of the Canary Islands (Fig. 5-6), where the water temperature was 20.1°C. The Type I genotype is clearly less common than the Type II genotypes, in the surface waters of the Northeast Atlantic during January.

5.3.2. Distribution of *Globigerinella siphonifera* Type II

A total of 39 *G. siphonifera* Type II specimens was sequenced, representing 28 specimens of Type IIa(3) and 11 specimens of Type IIb. Type IIa(3) was extremely widespread as it was found across the 3 water masses of the region, namely the North Atlantic Current water (NAC), the NATW and the Azores Current water (AC) (Fig. 5-6). The Type IIb genotype was less commonly found, although its distribution also stretched from the NAC water to the AC water (Fig. 5-6). The two

genotypes were found co-existing within the water column in the NAC, the AC and the Canary Basin.

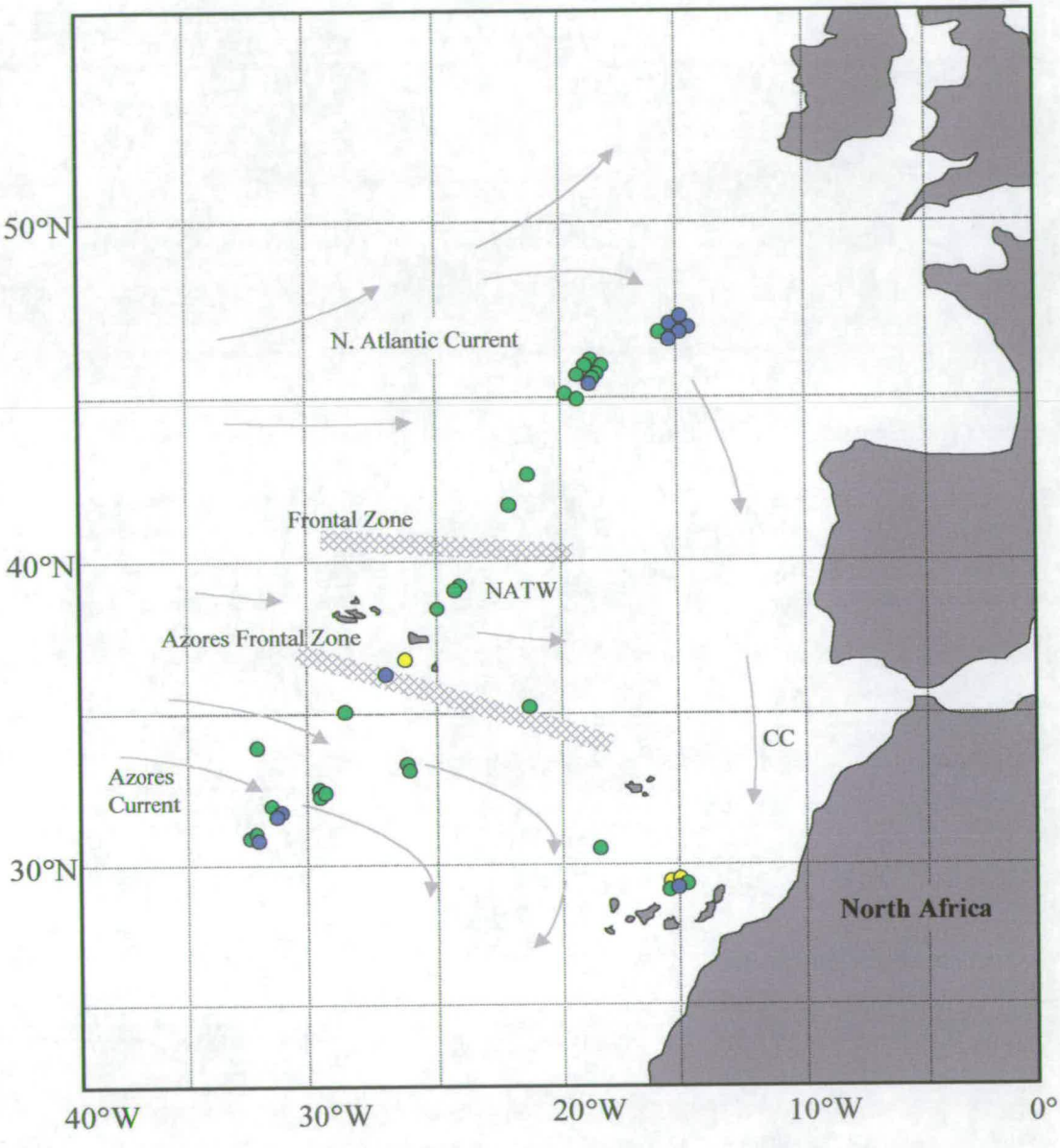


Fig 6. The distribution of *Globigerinella* genotypes within the North Atlantic sampled during collections M38/2 and P247. The genotypes are denoted: *G. siphonifera* Type I (○), *G. siphonifera* Type IIa(3) (●) and *G. siphonifera* Type IIb(●). The major surface currents are indicated with **CC** representing the Canary Current. The approximate location of the water mass frontal zones are shown, as determined by the shipboard thermo-salinometer. **NATW** denotes the North Atlantic Transitional Water. The 5 specimens located north of the Canary Islands, within the Canary Basin were collected during M38/2, with all of the other specimens collected during P247.

5.3.2. Distribution of *G. calida*

A total of 17 specimens of *G. calida* was sequenced and the genotype distribution is shown in Fig. 5-7. The majority of the *G. calida* specimens were found within the water of the NAC, although a single specimen was also found within the NATW. A further 2 specimens were found within the AC region, north-west of the Canary Islands (Fig. 5-7), which possessed a few genetic differences from the genotypes sequenced further north (see section 5.2.4).

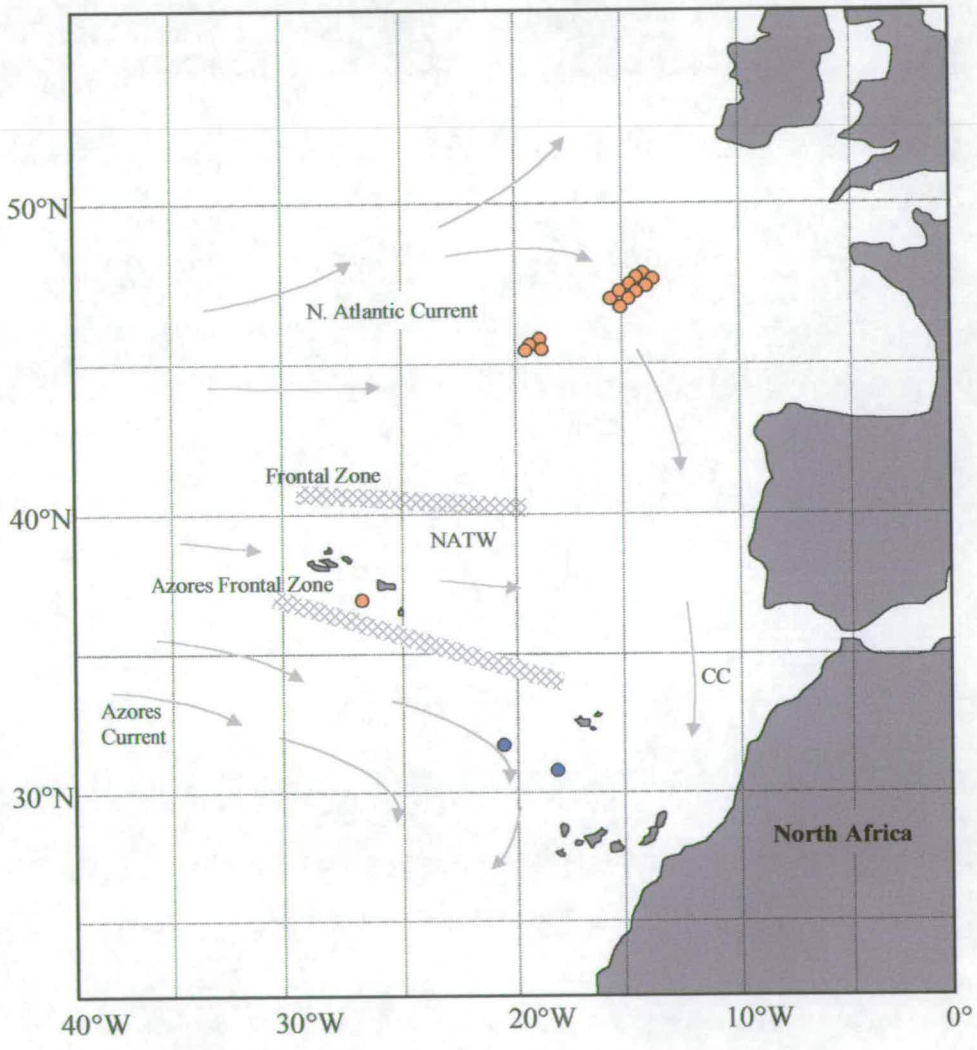


Fig. 5-7. Distribution of *G. calida* (●) within the North Atlantic. The variant genotypes (●) are also shown The major surface currents are indicated with **CC** denoting the cool Canary Current. The approximate location of the water mass frontal zones are shown, as determined by the shipboard thermo-salinometer. **NATW** denotes the North Atlantic Transitional Water.

5.4. Morphological variability within *Globigerinella* sp.

The identification of morphospecies within this genus was difficult due to the presence of many juvenile specimens, which have not reached their mature size or form. This is a particular problem for discriminating *G. siphonifera* from *G. calida*, since only the final few chambers make them recognisable from one another. In addition, it was not possible to separate *G. siphonifera* Type I and Type II. The specimens shown in Plate 5-1 are from bulk plankton samples, and illustrate the morphological variability within the *Globigerinella* genus.

Plate 5-1 description: *Globigerinella* sp. from the transitional-subtropical North Atlantic. The specimens were obtained from the NAC, NATW and the AC. The scale bars all represent 100μm, unless otherwise indicated. The scale bar for the magnified images of the test wall represents 20μm. **Fig. 1.** *G. siphonifera*, NAC specimen, trochospiral side. **Fig. 2.** *G. siphonifera*, NATW specimen, umbilical side. **Fig. 3.** *G. siphonifera/calida*, NATW specimen, umbilical side. **Fig. 4.** *G. siphonifera*, AF specimen, umbilical/side view. **Fig. 5.** *G. siphonifera*, AC specimen, umbilical side. **Fig. 6.** *G. siphonifera*, AC specimen, side view. **Fig. 7.** *G. siphonifera*, AC specimen, umbilical side. **Fig. 8.** *G. siphonifera*, AC specimen, trochospiral/side view. **Fig. 9.** *G. siphonifera*, AC specimen, trochospiral side. **Fig. 10.** Enlargement of terminal chamber, same specimen as shown in Fig. 3. **Fig. 11.** Enlargement of terminal chamber, same specimen as shown in Fig. 5. Note the large pore sizes compared with Figs. 10, 12 and 13. **Fig. 12.** Enlargement of terminal chamber, same specimen as shown in Fig. 7. Note the prominent spine bases. **Fig. 13.** Enlargement of terminal chamber, same specimen as shown in Fig. 8. Note the heavier calcification than in Fig. 12, and the prominent spine bases.

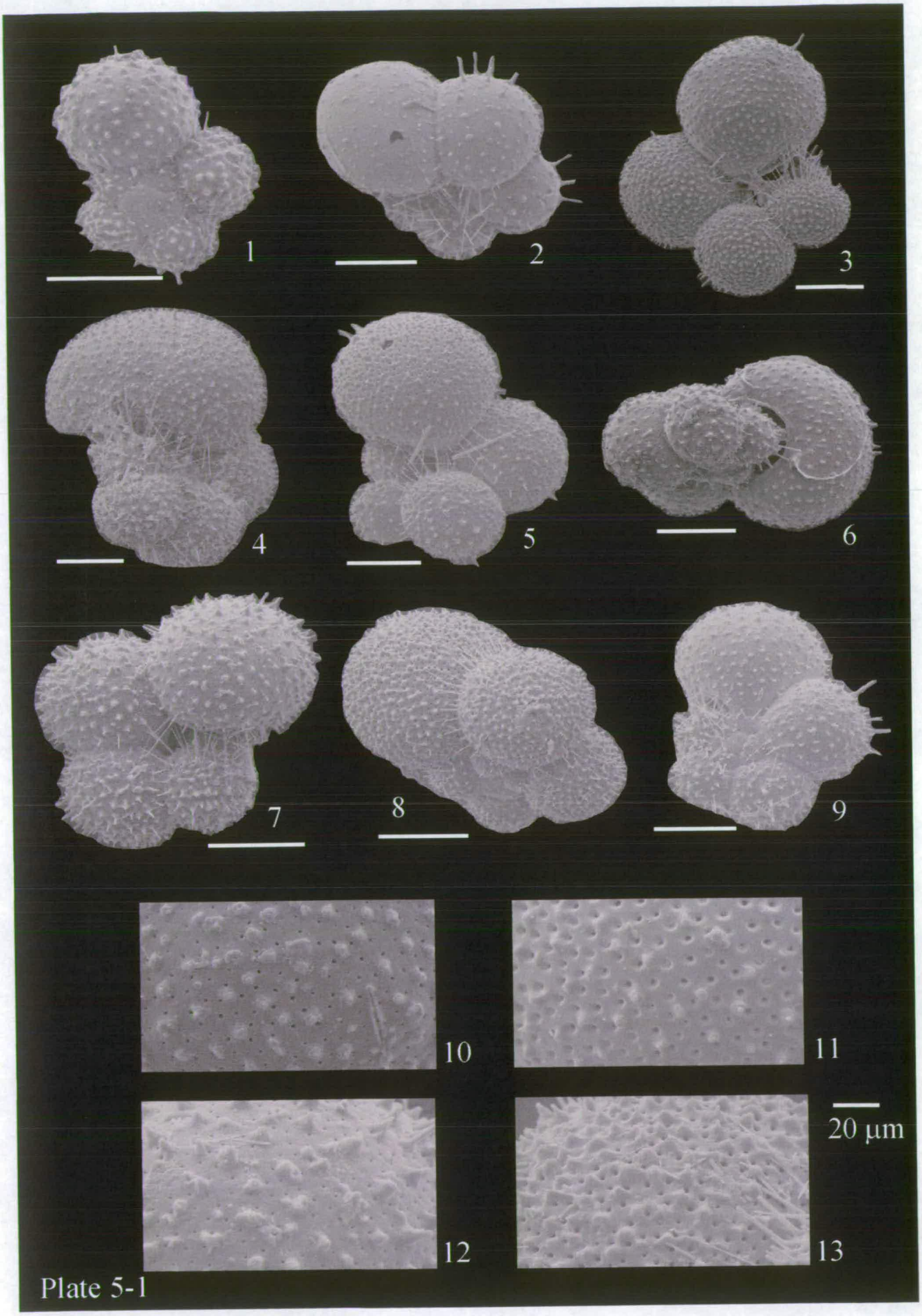


Plate 5-1

The coiling direction was noted for all specimens sequenced from the transitional-subtropical collection (from the Poseidon cruise). The coiling direction of the sequenced specimens of *G. siphonifera* Type IIa and IIb, and *G. calida* was ~ 50 % sinistral to dextral coiling. The *G. siphonifera* Type I specimen sequenced from the NATW was dextral coiling. The coiling direction of the *G. siphonifera* specimens from north of the Canary Islands is not known.

5.5. Discussion

5.5.1. Distribution of genotypes within the North Atlantic and their relationship to the ocean environment

Within the transitional-subtropical North Atlantic the *Globigerinella* species were found across the 3 major water mass regions. It is apparent that within the transitional-subtropical North Atlantic, *G. siphonifera* Types I and II have different distributions and levels of abundance (Fig. 5-6). The Type II genotypes are found extensively across the 3 main water mass regions of the Northeastern Atlantic sampled (Fig. 5-6). In contrast however, Type I was rare, with only 3 specimens being obtained. One specimen was obtained from the waters just north of the Azores frontal zone, whilst 2 specimens were obtained from the waters north of the Canary Islands (Fig. 5-6). In addition, the genotypes representing *G. calida* may also have differing distributions (Fig. 5-7). If these patterns are a result of habitat preferences, it could provide further proxies for examining the marine sediment record.

From the evidence of the differing abundances/distributions of the *G. siphonifera* genotypes, it is possible that *G. siphonifera* Types I and II have different

habitat preferences, although sampling density of Type I was low. During the Poseidon 247 collection (January), the sea surface temperature was at its lowest in the N. Atlantic (Levitus and Boyer, 1994). Primary production is relatively low in the North Atlantic during January, and is also quite patchy (Fig. 5-7). In this region during January, the highest primary productivity occurs in the NAC, but is patchy. Further south, the waters have relatively low primary production (Fig. 5-7).

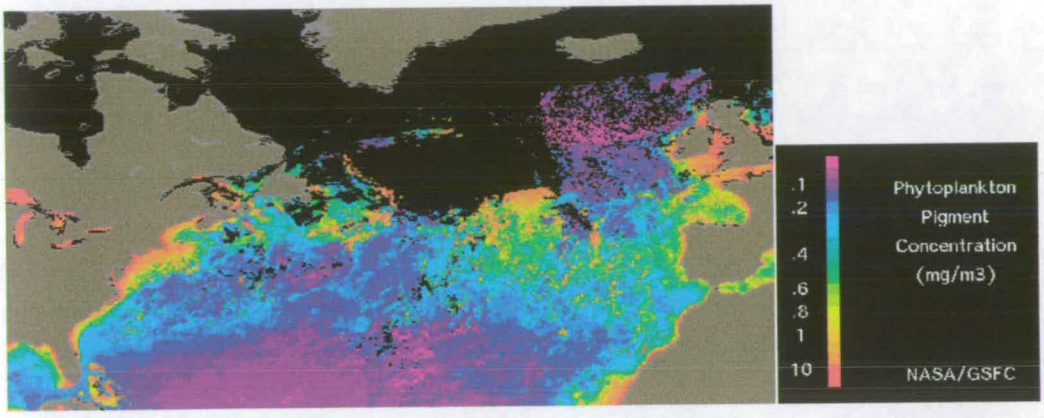


Fig. 5-7. Primary production within the North Atlantic during January (1997). From NASA/GSFC SeaWifs Ocean Colour website. The scale indicates the phytoplankton pigment concentration.

Therefore, the genotype distribution suggests that Type I may prefer warmer waters and/or different productivity conditions than the Type II genotypes. In contrast to the *G. siphonifera* genotype distribution/abundance found in the North Atlantic (this study), in the warm Caribbean Sea *G. siphonifera* is dominated by Type I, and Type II has a very low abundance (Bijma *et al.*, 1990a). This supports the suggestion that Type I prefers warmer waters than Type II. If the genotypes have different preferences for ocean productivity levels then this is not clear, since productivity levels in the North Atlantic were very low during the January collection (Fig. 5-7), and the Caribbean Sea is oligotrophic. It is possible that habitat tolerance

is reflected in the symbionts that each type possesses. Type I specimens have a relatively low number of facultative symbionts compared with the relatively high number of obligatory symbionts possessed by Type II specimens (Bijma *et al.*, 1998). This indicates that the potential photosynthetic rate of the Type II symbionts is higher than in Type I (Bijma *et al.*, 1998). However, it is unclear the effect this would have on the tolerance of the different types within the North Atlantic region. It could be supposed that the higher potential photosynthetic rate of Type II would make it better equipped to tolerate more variable oceanic conditions than Type I specimens. The differences in genotype distribution/abundance may simply be that Type I is specialised to warm oligotrophic waters and Type II is more of a generalist, capable of living in transitional/subtropical waters with variable productivity levels.

Analysis by Huber *et al.* (1997) and Bijma *et al.*, (1998) suggested that although Type I and Type II are found in sympatry within the surface waters, Type II tends to live deeper in the water column than Type I. The low numbers of *G. siphonifera* Type I obtained during the transitional-subtropical collections are therefore not likely to be explained by Type I living at a deeper level within the water column, thus evading capture during pump collection. Without further resampling, the possibility that the observed distribution patterns are a result of historical chance effects cannot be discounted. However, the distribution/abundance pattern is most likely explained by different spatial and temporal (seasonal) differences between Types I and II.

The *G. siphonifera* Type IIa(3) and Type IIb genotypes obtained from the transitional-subtropical N. Atlantic displayed similar extensive distributions, being

found across all three transitional-subtropical water masses of the region (Fig. 5-6). This suggests that they have similar habitat preferences.

Two genotypes of *G. calida* were also found within the North Atlantic. One genotype was obtained from the NAC and the NATW, whilst the other was found in the AC region (Fig. 5-7). It is possible that the variant genotype, found in the AC, prefers warmer water. However, due to the low specimen numbers collected, further work will be required to determine *G. calida* genotype distribution within the North Atlantic. Within the North Atlantic, *G. calida* has been identified within both the marine sediments (Funnell and Swallow, 1997) and the surface waters (this study) of the NAC. Therefore, it is clear that this morphospecies is not entirely restricted to subtropical/tropical water masses, but can also inhabit transitional water masses.

5.5.2. Morphology/genotype relationships

Although three *G. siphonifera* genotypes and two *G. calida* genotypes were identified within the transitional-subtropical North Atlantic, they could not be discriminated with the bulk plankton samples. The magnified views of the test wall reveal some differences between the samples, e.g. different pore sizes and the prominence of spine bases varies (Plate 5-1). Huber *et al.* (1997) noted that the morphological variation between *G. siphonifera* Type I and Type II was only found with detailed observations. Whether any morphological variation exists between the *G. siphonifera* Type IIa and Type IIb genotypes, and between the two *G. calida* genotypes, remains to be determined. In the case of *G. siphonifera* Types IIa and IIb, this is complicated further since they often co-exist within this region of the North

Atlantic. Resolving this issue will require combining genotyping with detailed morphological analysis, utilising a large number of mature specimens.

5.5.3. Genotypic variation within the *Globigerinella* clade

There is already good evidence that *G. siphonifera* Type I and Type II are distinct cryptic species (Darling *et al.*, 1997; Huber *et al.*, 1997; Bijma *et al.*, 1998). The molecular data presented in this study adds further support to the argument since Types I and II from the North Atlantic are shown to have a mean evolutionary distance of 7 % within the 767 bp phylogeny (Fig. 5-2). In addition to the molecular evidence, the genotype distribution pattern suggests that Types I and II may have different spatial and temporal (seasonal) abundances. This possibly has significant implications for using N. Atlantic *G. siphonifera* as a proxy within transfer functions, especially if Types I and II have different habitat preferences.

In addition to species level distinction between Type I and Type II, the molecular phylogeny (Fig. 5-3) has very strong bootstrap support for the division of genotypes within the Type II cluster (this study; Darling *et al.*, 1999). The Type II cluster subdivides into Type IIa and Type IIb, which are separated by a mean evolutionary distance of 2.1 % and supported in 98 % and 100 % of the bootstrap replicates respectively (Fig. 5-3). Within the North Atlantic collections, the Type IIa(3) and IIb genotypes showed similar distributions. The strong statistical support for their separation within the phylogeny, combined with the large number of sequence variations within the variable regions of the SSU rDNA fragment (Fig. 5-4), suggest that perhaps the North Atlantic *G. siphonifera* Type IIa(3) and Type IIb genotypes are also cryptic species. As the genotypes co-exist throughout most of the

study area, their genetic divergence must be maintained through reproductive isolation, which is the basis for biological speciation. Further, the low evolution rate characteristic of the *Globigerinella* lineage (see Chapter 3, section 3.4.2) suggests that the Type IIa and IIb genotypes may have been divergent for a considerable period of time. Further investigation will be required to determine whether the two genotypes have any associated morphological characteristics that would allow their discrimination. Any morphological variability between Type IIa and IIb is likely to be small, especially when the morphological differences between Type I and II are slight (Huber *et al.*, 1997). However, with additional work, this could provide a further proxy with which to examine the marine sediments.

Genetic variation was also found within North Atlantic *G. calida*. A total of 17 individuals was sequenced, and 2 individuals were found to differ from the others by a number of genetic differences. Unfortunately, as only a 500 bp SSU rDNA fragment was amplified for the variant genotype, the significance of the sequence variations remains to be determined and additional work will be required to resolve whether the genotypes display distinct distribution patterns.

5.5.4. Gene flow within the North Atlantic, Caribbean and Mediterranean region

Each of the *G. siphonifera* genotypes obtained from the North Atlantic contributes towards understanding gene flow within the World's oceans. The *G. siphonifera* Type I genotype from the North Atlantic is identical to the Type I genotypes from the Caribbean (Fig. 5-1), indicating that there must have been recent genetic mixing between the populations, or perhaps that it occurs constantly *via* the

N. Atlantic gyre. It is possible that *G. siphonifera* Type I is represented by a single continuous genotype population in this region. In contrast, the North Atlantic Type IIa(3) and Caribbean Type IIa(1) genotypes have a number of sequence differences indicating that they do not mix as frequently as the Type I genotypes from these regions. Due to the prevailing surface currents these regions are, at present, not isolated from each other. However, it is possible that the genotypes became isolated during glacial/interglacial cycling for a sufficient time to produce the sequence divergence. The isolation may have arisen due to the populations being separated by a stretch of unfavourable ocean habitat caused by the climatic cycling, such that the populations were unable to survive transit and could not mix genetically. This contrasts with the homogeneity of the Type I genotype and suggests that perhaps Type I maintained a single population in the North Atlantic region during the glacial/interglacial cycling (perhaps by retreating to warmer tropical waters). After the end of the last glacial period, the Type I genotype would have been free to propagate into the areas where we find it today. It is unknown whether the Type IIa(3) and the Type IIa(1) genotypes exist in both the Caribbean and the North Atlantic. The North Atlantic *G. siphonifera* Type IIa(3) genotype is identical to the Mediterranean specimen sequenced by de Vargas *et al.*, (1997), indicating that genetic interchange occurs between the N. Atlantic and Mediterranean. This supports the *G. bulloides* Type Ib data (Chapter 4, section 4.5.6) that suggested that the direction of gene exchange would most likely be from the North Atlantic to the Mediterranean due to the flow direction of the surface currents. The other *G. siphonifera* Type II genotypes are considered in Chapter 8, where gene flow is discussed on a global scale.

Chapter 6: *Globigerinoides ruber* (d'Orbigny)

Index

6.1. Introduction

In this chapter, genotypes of *Globigerinoides ruber* (d'Orbigny) obtained during Meteor cruises M37/2 and M38/2 are examined. An introduction to the morphospecies within the *G. ruber/conglobatus* cluster and their distribution is followed by an analysis of the molecular data obtained during this study. The chapter concludes with a discussion of the results obtained.

Globigerinoides ruber has two forms: a pink form and a white form. The colouration of the pink form is thought to be due to the presence of a pigment called phaeophytin (Bé and Tolderlund, 1971). It had been suggested that the pink and white forms may be distinct subspecies (Deuser and Ross, 1989; Pujol and Grazzini, 1995) and their species status was confirmed by Darling *et al.* (1999) on the basis of molecular phylogenetic evidence. Darling *et al.* (1999) suggested that the pink and white forms may have diverged as long as 11 million years ago. This may have gone unseen in the fossil record since the pink form may not have been pink when it diverged (Darling *et al.*, 1999), or the pink pigment may have faded in the sediments over such a long period of time (Thompson *et al.*, 1979). It was also shown that there are two extant *G. ruber* lineages which may have diverged as long as 22 million years ago. One lineage comprises *G. ruber* pink and white, the other comprises *G. ruber* white and *G. conglobatus* (Darling *et al.*, 1999).

The stratigraphic distribution of *G. ruber* is somewhat confused by varying interpretations of the evolution of the lineage, although it is thought that *G. ruber* has been in existence since the Middle Miocene (Zone N16) (Kennett and Srinivasan, 1983; Blow, 1969; Cordey, 1967). However, in light of the molecular phylogenetic

evidence, Darling *et al.* (1999) suggested that the two *G. ruber* lineages possibly diverged during the early Miocene.

If the *G. ruber* morphotypes are considered collectively, they are found from transitional to tropical water masses, across a temperature range of 14-30°C (Bé and Tolderlund, 1971). In an attempt to determine the tolerance of *G. ruber* to varying temperature and salinity, Bijma *et al.* (1990a) carried out a series of culture experiments. The study had particular relevance to the distribution of this morphospecies within the oceans, although the pink and white forms were not differentiated. It was found that *G. ruber* could tolerate a temperature range of 14-32°C which is very similar to that found within its natural habitat, and that it could tolerate a salinity range of 22-49 ‰. The specimens of *G. ruber* did not live as long in culture as any of the other species studied (*G. sacculifer* and *G. siphonifera*), which lead Bijma *et al.* (1990a) to conclude that *G. ruber* has a different food requirement from the other species.

It is known that the pink and white forms have markedly different distribution patterns. Especially important is that the pink form became extinct within the Indo-Pacific during the Late Pleistocene, ~ 120, 000 years ago (Thompson *et al.*, 1969), with only the white form persisting. This provides an excellent datum for Quaternary marine sediments, and has implications for gene flow in the oceans since the pink form has not repopulated the Indo-Pacific from the Atlantic. The extinction of the pink form from the Indo-pacific is also evidence that the pink and white forms are distinct species. Both pink and white forms have persisted within the Atlantic and Mediterranean, although they have quite different spatial distributions, with the white form being generally more common throughout this region. The pink form is less

common in the South Atlantic compared with the North Atlantic, and is predominantly found in waters near land masses (Bé and Tolderlund, 1971). Within the waters of the eastern North Atlantic, Ottens (1992) found *G. ruber* pink in August south of 33°N. Indeed, Bé and Hamlin (1967) found that the pink form was very abundant in the western Sargasso Sea, but was sparse in the central and eastern North Atlantic. In the sediments of the eastern N. Atlantic the northern-most occurrence of the white and pink form is 45°N and 39°N, respectively (Ganssen and Kroon, in press).

As well as a spatial difference in distribution patterns between the pink and white forms of *G. ruber*, there have been accounts of a temporal difference also. While *G. ruber* white is perennially abundant within the Sargasso Sea (Deuser and Ross, 1989), *G. ruber* pink is most common (highest proportion of total assemblage) during the summer, at the time of elevated temperature when the surface waters are well stratified. During this period, the white form is at a reduced level (Deuser and Ross, 1989; Pujol and Vergnand-Grazzini, 1995).

Within the water column *G. ruber* is a shallow dweller, living predominantly in the upper 50 m (Bé, 1977). This is in agreement with isotope studies by Deuser and Ross (1989) and Ganssen and Kroon (in press), which showed that both pink and white forms lived very near the surface. Hemleben *et al.* (1989) showed that associated symbionts are dinoflagellates, similar to those found in other *Globigerinoides* species and *Orbulina universa*. This was subsequently confirmed by molecular phylogenetic analysis where it was shown that the symbionts were *Gymnodinium beii* (Gast and Caron, 1996).

With such a wide tolerance to temperature and salinity it is no surprise that *G. ruber* is the most abundant planktic foraminifer in subtropical waters. This morphospecies can account for up to 50 % of the total foraminiferal population in the Canaries Current, North Atlantic Current, Caribbean, Antilles Current and the Gulf Stream (Bé and Tolderlund, 1971). In the western North Atlantic, Cifelli and Smith (1970) found that *G. ruber* could account for up to 71 % of the foraminiferal assemblage, with Ottens (1991) showing that in the eastern North Atlantic *G. ruber* dominated the subtropical assemblage of the Azores Current.

Globigerinoides ruber is a potentially important source of information for paleoceanographic reconstructions, mostly due to its shallow dwelling nature. Isotope studies allow a reconstruction of sea surface temperatures, with *G. ruber* pink in particular reflecting summer sea surface temperatures (Ganssen and Kroon, in press). In addition, being a major component of the assemblage, *G. ruber* is important to sea surface temperature estimates using transfer functions. Due to the differing distribution patterns of the pink and white forms, they should be differentiated wherever possible since they reflect different ocean conditions.

6.2. Evolutionary relationships within the *G. ruber/conglobatus* clade

A total of nine *G. ruber* specimens from the Canary Basin, collected during cruises M37/2 and M38/2 from the time-series station ESTOC (European Station for Time Series Observations, Canary Islands), were sequenced (Fig. 6-1).

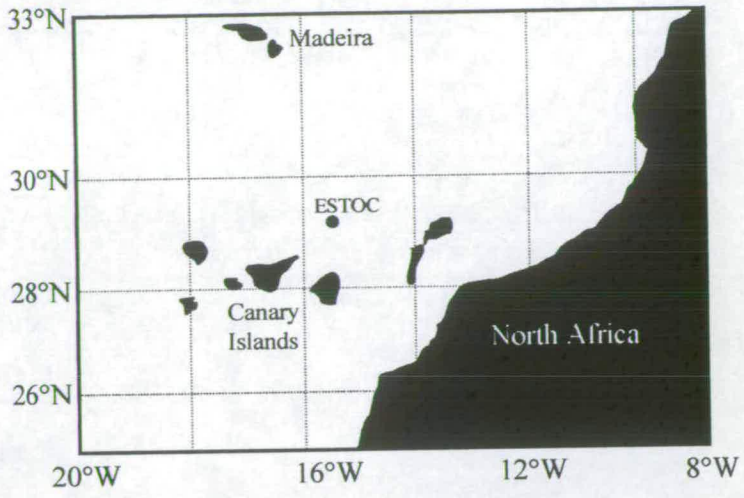


Fig. 6-1. Location of ESTOC, north of the Canary Islands, where the *G. ruber* specimens were obtained.

The neighbour-joining phylogeny presented in Fig. 6-2 represents a sub-tree of the more extensive phylogeny based on 505 unambiguously aligned sites discussed in Chapter 3. The tree illustrates the phylogenetic position of the three *G. ruber* genotypes collected from north of the Canary Islands, in the North Atlantic, during Meteor cruises M37/2 and M38/2. Unfortunately, due to degradation of DNA no *G. ruber* partial SSU rDNA sequences were obtained from the Poseidon 247 collection.

The *G. ruber/conglobatus* lineage divides into two main clusters (Fig. 6-2), both of which are supported in 100 % of the bootstrap replicates. The two clusters are separated by a mean evolutionary distance of 10.1 % in the 505 bp molecular phylogeny. For ease of description the clusters have been defined as I and II (Fig. 6-2).

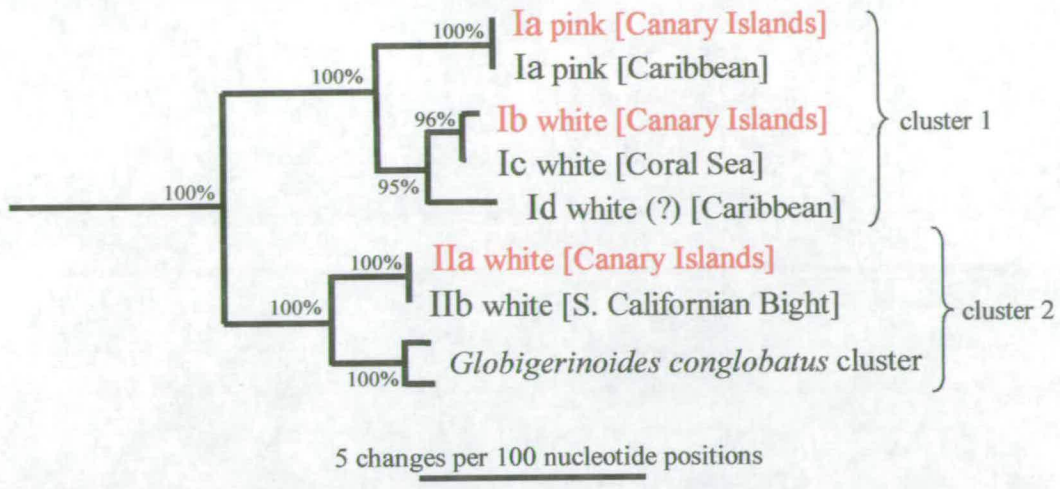


Fig. 6-2. Sub-tree of neighbour-joining phylogeny based on 505 unambiguously aligned sites, illustrating the evolutionary relationships within the *Globigerinoides ruber/conglobatus* clade. Support has been assigned to the phylogeny using 1000 bootstrap replicates. The test colouration of Type Id (Pawlowski *et al.*, 1997) is not known, but it was most probably white as it clusters with the other white form genotypes.

6.2.1. Type I cluster

The Type I cluster has support from 100 % of the bootstrap replicates (Fig. 6-2). The branch to this cluster divides into two, representing the division between *G. ruber* pink (Type Ia) and *G. ruber* white (Type Ib, Ic and Id). The *G. ruber* pink and *G. ruber* white clusters are supported in 100 % and 95 % of the bootstrap replicates respectively. The *G. ruber* pink genotypes from the Canary Basin and the Caribbean are identical throughout the entire ~1000 bp SSU rDNA fragment (Fig. 6-3). The *G. ruber* pink genotype is separated from the *G. ruber* white genotypes of the Canary Basin (Type Ib), the Coral Sea (Type Ic) and the Caribbean (Type Id) by evolutionary distances of 5.2 %, 4.8 % and 5.0 % respectively. Comparison of the *G. ruber* pink (Canary Basin) genotype with the *G. ruber* white Type Ib (Canary Basin) genotype shows that there is considerable sequence divergence in the variable

regions of the SSU rDNA fragment (Fig. 6-3), with the foraminiferal specific insertion, F3, and the variable region, V9, being unalignable (Fig. 6-3).

The *G. ruber* white Type I cluster divides again, representing the division between the Canary Basin (Type Ib) and Coral Sea (Type Ic) genotypes, and the Caribbean (Id) genotype. The Canary Basin and Coral Sea genotype cluster has support from 96 % of the bootstrap replicates. The Canary Basin (Type Ib) genotype is separated from the Coral Sea (Type Ic) genotype and the Caribbean (Type Id) genotype by evolutionary distances of 0.4 % and 2.4 % respectively. Comparison of the Canary Basin (Type Ib) and the Coral Sea (Type Ic) genotype nucleotide sequences shows that they are actually separated by only 2 base substitutions across the entire ~1000 bp SSU rDNA fragment which are located in the conserved region, C7 (Fig. 6-3). Comparison of the Canary Basin (Type Ib) and the Caribbean (Id) genotypes shows that there are 29 base substitutions and 8 deletion/insertion differences throughout the entire ~ 1000 bp SSU rDNA fragment (Fig. 6-3).

6.2.2. Type II cluster

The Type II cluster has support from 100 % of the bootstrap replicates (Fig. 6-2). The branch to this cluster divides into two, representing the division between the *G. ruber* white (Type IIa and Type IIb) and the *G. conglobatus* clusters. Each cluster has support from 100 % of the bootstrap replicates. The *G. ruber* white genotypes are separated from the *G. conglobatus* genotypes by a mean evolutionary distance of 3.9 %. Within the 505 bp phylogeny the Canary Basin (Type IIa) and the S. Californian Bight (Type IIb) genotypes appear to be identical, although comparison of the

genotype SSU rDNA sequences shows 1 substitutional base difference within the foraminiferal specific insertion, F2 (Fig. 6-3).

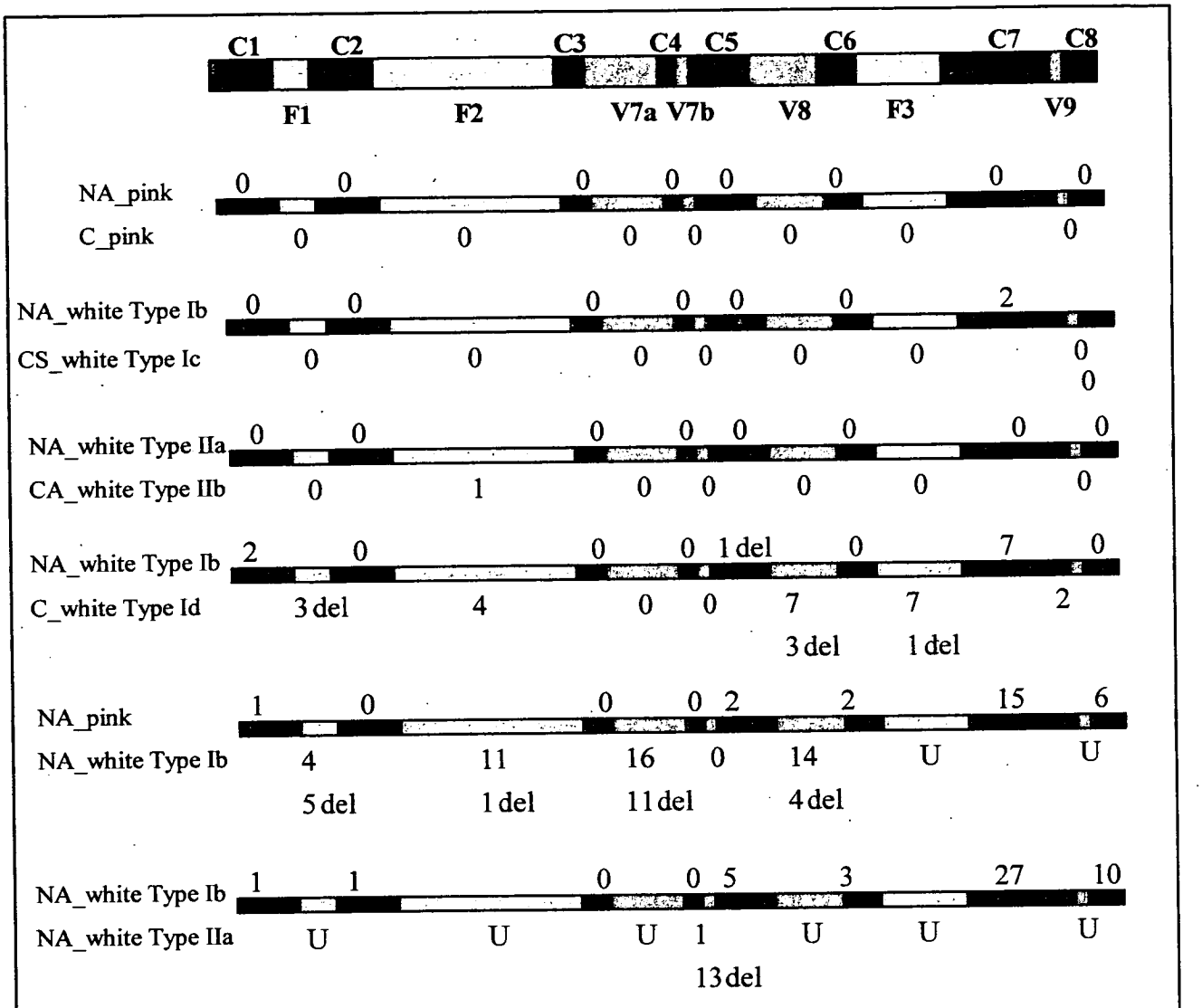


Fig. 6-3. Schematic representation of the ~1000bp 3' terminal region of the SSU rRNA gene. Comparisons of the genetic differences are made between the genotypes from the North Atlantic, Southern Californian Bight, Caribbean, and Coral Sea. The number of base differences between each of the genotypes are indicated, with the number of base substitutions and sequence length variations (base deletions/insertions) shown. **C1-C8** represent the highly conserved regions which were aligned relative to comparable regions present in all eukaryotes, **V7-V9** represent variable length expansion segments present in most eukaryotes and **F1-F3** represent three insertions which are unique to foraminifera. **bs** represents base substitutions, **del** represents sequence length variations due to insertions/deletions and **U** denotes an unalignable region. **NA** denotes the Canary Basin, **C** denotes the Caribbean, **CA** denotes the S. Californian Bight, and **CS** denotes the Coral Sea.

6.2.3. Phylogenetic comparison of the Canary Islands and Caribbean *G. ruber* genotypes

Four specimens of *G. ruber* pink were sequenced and all were found to be identical. Five specimens of *G. ruber* white were sequenced, and were found to consist of two different genotypes; a single *G. ruber* white Type I specimen and four specimens of the *G. ruber* white Type II variety. The three genotypes obtained from the Canary Islands collections are all quite divergent from one another (Fig. 6-2). The *G. ruber* pink genotype is separated from the *G. ruber* white Type Ib and Type IIa genotypes by evolutionary distances of 5.2 % and 10.2 % respectively. Further, the *G. ruber* white Type Ib and Type IIa genotypes are separated by an evolutionary distance of 9.4 %. Comparison of the *G. ruber* white Type Ib and Type IIa genotypes from the Canary Basin show that in addition to the 47 base substitution differences within the conserved regions (Fig. 6-3) of the SSU rDNA fragment, the variable regions have high sequence divergence since all but one region is unalignable (Fig. 6-3). The three genotypes also have different sequence lengths for the amplified region: Type Ia, Ib, and IIa have sequence lengths of 988 bp, 975 bp, and 995 bp respectively.

If we consider that Pawlowski *et al.* (1997) also sequenced a *G. ruber* white Type Id genotype from the Caribbean, which is separated from *G. ruber* white Type Ib by an evolutionary distance of 2.4 %, there are at least three distinct *G. ruber* white genotypes existing within the North Atlantic/Caribbean region. In total, therefore, at least four distinct *G. ruber* genotypes exist within the North Atlantic/Caribbean region.

6.3. Morphology of *G. ruber*

Although no specimens of *G. ruber* were collected for morphological analysis during either of the Meteor cruises (M37/2 and M38/2), specimens were collected in bulk plankton collections during the other transects and are shown in Plate 6-1. One of the *G. ruber* white specimens was obtained from the subarctic (60°00.4N/14°32.3W-59°49.2N/19°11.6W). This highlights the potential for planktic foraminifers to become expatriated as they are passively carried in ocean currents, into regions where they are not adapted to live. The other *G. ruber* white specimens were collected from the NATW, the Azores frontal zone, and the AC (Fig. 2-3).

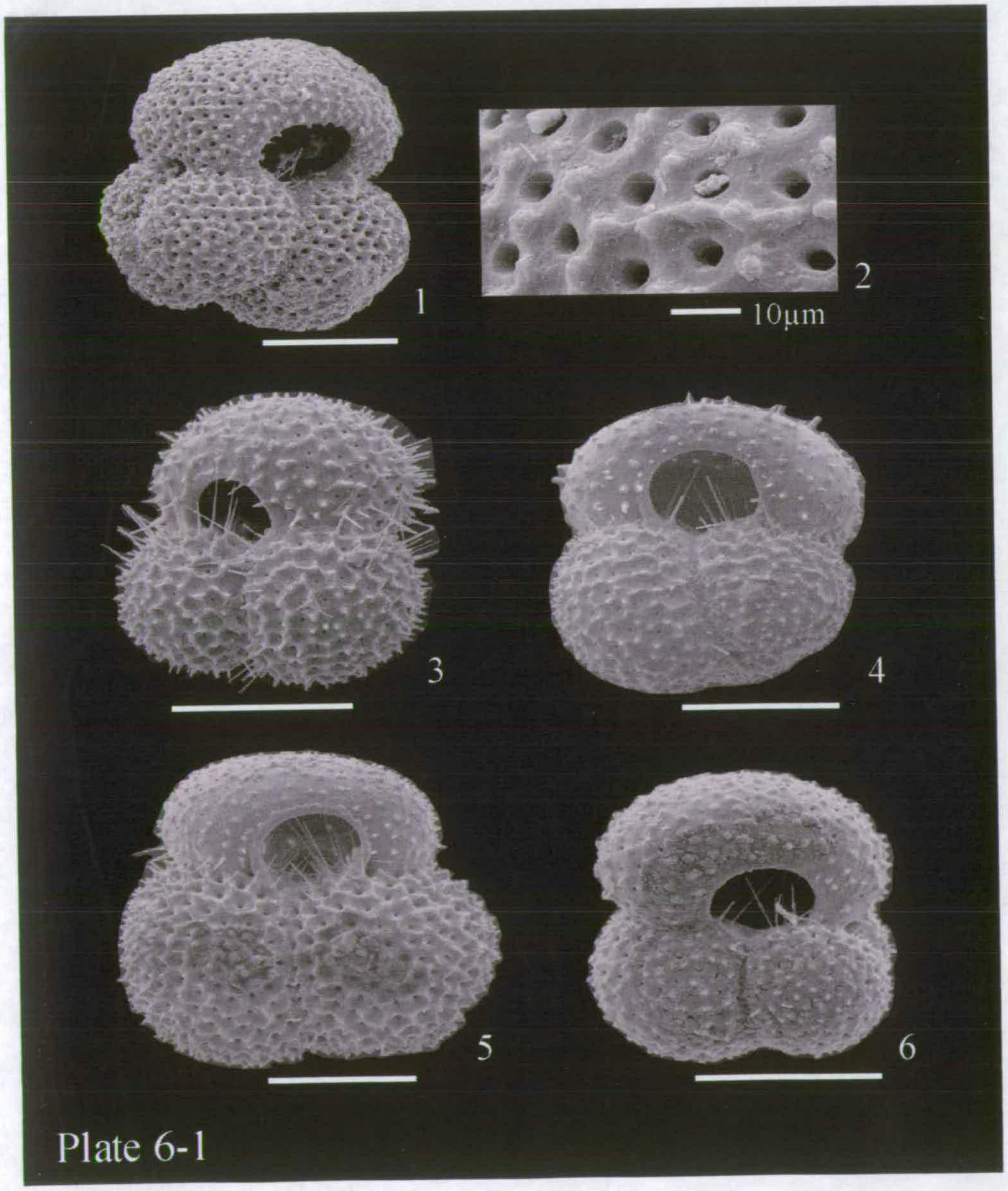


Plate 6-1 description: *Globigerinoides ruber* white specimens from the North Atlantic. The scale bars represent 100μm unless otherwise indicated.

Fig. 1: subarctic specimen, mature, umbilical side. **Fig. 2:** enlargement of terminal chamber of specimen in Fig. 1. Note the large pore sizes compared with *G. bulloides* (Plates 4-1 and 4-2) and *G. siphonifera* (Plate 5-1). **Fig. 3:** North Atlantic Transition Water specimen, juvenile, umbilical side. **Fig. 4:** Azores frontal zone, juvenile, umbilical side. **Fig. 5:** Azores frontal zone, umbilical side. Note the diminutive terminal chamber compared with the terminal chamber in Fig. 4, and heavy calcification of 2nd and 3rd last chambers. **Fig. 6:** Azores Current specimen, juvenile, umbilical side.

6.4. Discussion

6.4.1. Genotypic variation within North Atlantic *Globigerinoides ruber*

It has been shown that individual spinose planktic foraminiferal morphospecies represent a cluster of several genotypes (Darling *et al.*, 1999; this study). The results within this chapter show that at least four *G. ruber* genotypes inhabit the North Atlantic region. Three distinct genotypes have been obtained from the Canary Basin region of the North Atlantic, which were found to co-exist within the water column. As sampling numbers of *G. ruber* were very low, this must indicate that the genotypes represent significant proportions of the standing population.

The three North Atlantic genotypes sequenced in this study are separated by considerable evolutionary distances, and using the calibrated molecular phylogeny of

Darling *et al.* (1999), ancient divergences between the genotypes are indicated. Darling *et al.* (1999) suggest that the divergence between the *G. ruber* pink (Type Ia) and white (Type Ib) genotypes and the *G. ruber* white (Type IIa) genotype could have been as long ago as 22 Ma, in the early Miocene. Further, the divergence between the *G. ruber* pink (Type Ia) genotype and the *G. ruber* white (Type Ib) genotype was estimated at approximately 11 Ma. Even assuming a large error range, the estimated divergences point to ancient speciation events.

It seems quite likely that the three N. Atlantic *G. ruber* genotypes are distinct species, and therefore it is possible that each genotype may have different habitat preferences. The genetic data shows that the pink form is obviously a distinct species from the white forms, and it is known that they have different temporal and spatial distribution patterns (Deuser and Ross, 1989; Pujol *et al.*, 1995), therefore, any further work that requires the use *G. ruber* in the Atlantic Ocean or Mediterranean Sea should differentiate both types, or else error will possibly be introduced into investigations.

In addition, the two *G. ruber* white genotypes may have different habitat preferences. The molecular phylogeny (Fig. 6-2) shows that *G. ruber* white Type Ib (Canary Basin) clusters with *G. ruber* white Type Ic (Coral Sea) and Type Id (Caribbean), suggestive that Type I may have a preference for subtropical/tropical water. *G. ruber* white Type IIa (Canary Basin) clusters with *G. ruber* white Type IIb (S. California Bight), which suggests that Type II may prefer upwelling or transitional water. If these are the preferred habitats of the *G. ruber* white genotypes, not distinguishing them could affect sea surface temperature estimates based on transfer functions. Such potential adaptations warrant further investigation. As the

distinct genotypes are likely to have diverged millions of years ago, it is possible that some morphological changes may have occurred which could be correlated to genotype. If genotype could be linked to morphotype, it may provide further proxies with which to examine the deep-sea marine sediments. Indeed, if the distinct types do have different habitat preferences, it may possibly provide new proxies for upwelling/transitional or subtropical/tropical water masses.

6.4.2. Gene flow within the North Atlantic and Caribbean region

The molecular phylogeny (Fig. 6-2) shows that the *G. ruber* pink (Type Ia) genotype from the Canary Islands is identical to the genotype obtained from the Caribbean. Indeed, the two partial SSU rDNA sequences are identical throughout the entire ~1000 bp region (Fig. 6-3). This indicates that the two specimens are either part of a continuous population, or have mixed relatively recently, especially since there is not even a single base difference within the variable regions of the SSU rDNA fragment. This is understandable since the surface water currents within the North Atlantic gyral system could allow genetic exchange to occur with the Caribbean population, and *vice versa*. It is therefore probable that *G. ruber* pink found in other regions of the North Atlantic is the same genotype. In contrast, the *G. ruber* white Type Ib (North Atlantic) and Type Id (Caribbean) have an evolutionary distance of 2.4 % in the 505 bp phylogeny, and a considerable number of sequence differences in the variable regions of the SSU rDNA fragment (Fig. 6-3). This indicates that the genotypes have not mixed for a considerable period of time. Further investigation of the distribution of *G. ruber* white Type I genotypes is

required, to determine whether Type Ib exists in the Caribbean, and whether Type Id exists in the North Atlantic. It is possible that Type Id is specific to the Caribbean.

Chapter 7: Non-spinose planktic foraminifera

Index

7.1. Introduction

In this chapter, the non-spinose planktic foraminiferal species *Neogloboquadrina pachyderma* and *Globigerinita uvula* sampled in the North Atlantic are examined. The non-spinose planktic foraminifera form an important part of the total planktic foraminiferal assemblage around the World's oceans (Bé and Tolderlund, 1971; Bé, 1977), and a number of them provide proxies for past oceanographic/climatic change in the deep-sea sediment record. Initial interest in the non-spinose species was fuelled by the observation of sinistral and dextral coiling provinces, which could be used to correlate stratigraphic zones in deep-sea sediment cores (Ericson *et al.*, 1954; Ericson and Wollin, 1956). The importance of coiling direction ratios within *Neogloboquadrina pachyderma* was recognised (Ericson, 1959; Bandy, 1960), and currently *N. pachyderma* is perhaps one of the most intensively utilised planktic foraminiferal species for palaeoceanographic, palaeoclimatic and stratigraphic investigations.

Molecular phylogenetic analysis by Darling *et al.* (1997) has shown that the planktic foraminifera have polyphyletic origins. The non-spinose foraminifers cluster separately from the spinose species, falling amongst the benthic foraminifera sequenced by Pawlowski *et al.* (1997). Darling *et al.* (1999) concluded that the planktic foraminifers have different benthic ancestors and that translocation from the benthos to the plankton was not a single event.

The chapter starts with an introduction to *N. pachyderma* and *Globigerinita uvula*, giving details of their present day distribution and a review of important past research. The *Neogloboquadrina* clade is investigated, and the phylogenetic

placement, distribution and morphological variability of subarctic genotypes is described. The chapter concludes with a discussion of the results.

7.1.1. *Neogloboquadrina pachyderma* (Ehrenberg)

At present, *N. pachyderma* lives across a surface temperature range of between 0°C and 24°C, with a peak abundance between 0°C - 9°C (Bé and Tolderlund, 1971). Its diet consists of phytoplankton and is often diatoms (Hemleben *et al.*, 1989). Two forms of *N. pachyderma* are recognised: sinistral coiling (S) and dextral coiling (D) individuals. They have markedly different distributions, with the sinistral coiling form being a predominantly polar water inhabitant and the dextral coiling form preferring to live in subpolar to transitional waters (Bé and Tolderlund, 1971; Bé, 1977). Both sinistral and dextral coiling forms are also known to inhabit low latitude upwelling systems, with each having a different temporal and spatial distribution around these systems (Ufkes and Zachariasse, 1993; Ufkes *et al.*, 1998; Ivanova *et al.*, 1999).

The coiling direction of *N. pachyderma* has been traditionally used as an indicator of shifts in the polar / subpolar ocean isotherms, with the sinistral form indicating polar waters (Ericson, 1959; Bandy, 1960; Bé and Hamlin, 1967; Tolderlund and Bé, 1971). The presence of *N. pachyderma* (S) within the planktic foraminiferal assemblage has also been used as an upwelling indicator within the fossil record (Naidu and Malmgren, 1996a). However, the significance of temperature to *N. pachyderma* coiling direction has been questioned, since plankton collections by Cifelli (1971) in the North Atlantic's Labrador Current, where the water temperature was 1.9°C, yielded no sinistral coiling specimens. It was noted that

a dominantly (or entirely) sinistral coiling population should have characterised this region, if coiling direction is to be attributed to ocean temperature (Cifelli, 1971). More recent research has shown that in addition to sea surface temperature, nutrients and the availability of appropriate food must also be considered (Sautter and Thunell, 1991a; Ufkes and Zachariasse, 1993; Carstens and Wefer, 1992). Recent investigation of the morphological evolution of *N. pachyderma* (S) suggests that it may not have been polar adapted prior to approximately 1 Ma in the North Atlantic, with maximum polar adaptation occurring after 0.4 Ma (Baumann *et al.*, 1998; Huber *et al.*, 1998). Indeed, it appears that utilising *N. pachyderma* (S) older than 1 Ma for reconstructing high latitude sea surface temperatures, using transfer functions, would contain significant error (Huber *et al.*, 1998).

In addition to the discussion of the relative merits of coiling direction within *N. pachyderma* to palaeoceanography, there has also been debate over the taxonomy of the sinistral and dextral coiling forms (e.g. Cifelli, 1961, 1973). In the *neogloboquadrina* cline there is a morphological gradation from *N. pachyderma* (S) to *N. pachyderma* (D) to *Neogloboquadrina dutertrei* (Bé and Tolderlund, 1971). Recent investigation of the ecological preference of *N. pachyderma, N. dutertrei*, and morphologically intermediate forms (*N. pachyderma – N. dutertrei* intergrades) by Hilbrecht (1997) suggested that there are two major ecological groups in *Neogloboquadrina*: (1) *N. dutertrei* (tropical population); (2) *N. pachyderma* and *N. pachyderma – N. dutertrei* intergrades. Further, Hilbrecht (1997) suggested that subtropical *N. dutertrei* had a closer ecological relationship to the second group, such that it should not be lumped with tropical *N. dutertrei*.

Although Brummer and Kroon (1988) suggested that coiling direction might be a genetic trait, prior to the advent of molecular phylogenetic evidence it was unclear whether or not the sinistral and dextral coiling forms were distinct species. However, Darling *et al.* (submitted) has confirmed that within Antarctic *N. pachyderma* specimens, coiling direction is associated with genetically distinct populations and is not a phenotypic effect triggered by temperature. Further, it was shown that the dextral coiling forms are not an intermediate between *N. pachyderma* (S) and *N. dutertrei* (Darling *et al.*, submitted).

7.1.2. *Globigerinita uvula* (Ehrenberg)

Globigerinita uvula is a common species within high latitude assemblages (Parker, 1962; Saito *et al.*, 1981; Hemleben *et al.*, 1989). This morphospecies has not been used for paleoceanographic purposes.

7.2. Phylogenetic placement of the North Atlantic non-spinose planktic foraminifera *Neogloboquadrina pachyderma* and *Globigerinita uvula*

A total of 24 specimens of *N. pachyderma* (D) was sequenced from the subarctic Atlantic. A total of six specimens of *G. uvula* was sequenced from the North Atlantic. Partial (~1000 bp) SSU rDNA sequences have been obtained for both of these morphospecies and are included within the 505 bp phylogeny discussed in Chapter 3. Within the benthic and non-spinose planktic region of the 505 bp foraminiferal molecular phylogeny (Fig. 3-2), *N. pachyderma* (D) clusters with *N. dutertrei*, and *G. uvula* clusters near to *G. glutinata*.

Unfortunately, only 1 individual specimen of *N. pachyderma* (S) was successfully amplified and sequenced as very few specimens were collected. Only a ~500 bp SSU rDNA sequence was obtained from this specimen, in addition to similar sized fragments obtained from *Globorotalia inflata* (d'Orbigny) from the subarctic Atlantic, and *Neogloboquadrina dutertrei* (d'Orbigny) from the Azores Current region of the North Atlantic. The full ~1000 bp SSU rDNA fragment was not obtained due to problems in amplification. Since these sequences are only half the length required for inclusion within the main 505 bp phylogeny (Chapter 3), a phylogeny was constructed using a reduced number of base comparisons to determine their genetic relationships.

7.2.1. The *Neogloboquadrina* clade

In an attempt to resolve the relationships within the *Neogloboquadrina* clade, a neighbour-joining phylogeny was constructed using 284 unambiguously aligned sites (Fig. 7-1). The alignment and distance matrix for this phylogeny is shown in the Appendices (A2.1.5 and A2.2.5). Included within the phylogeny are the planktic foraminifers *N. pachyderma* (D) (subarctic, this study), *N. pachyderma* (S) (subarctic, this study), *N. dutertrei* (Azores Current region of North Atlantic, this study, see Fig. 2-3), *G. inflata* (subarctic, this study), *N. pachyderma*, dextral (subantarctic, Darling *et al.*, submitted), *N. dutertrei* (Caribbean, Darling *et al.* 1997) and *G. inflata* (Mediterranean, de Vargas *et al.*, 1997, GenBank Accession number Z83971). The *G. inflata* genotypes were used as an outgroup since the molecular phylogeny of de Vargas *et al.* (1997) suggested that it diverged from a common ancestor prior to *N. dutertrei*.

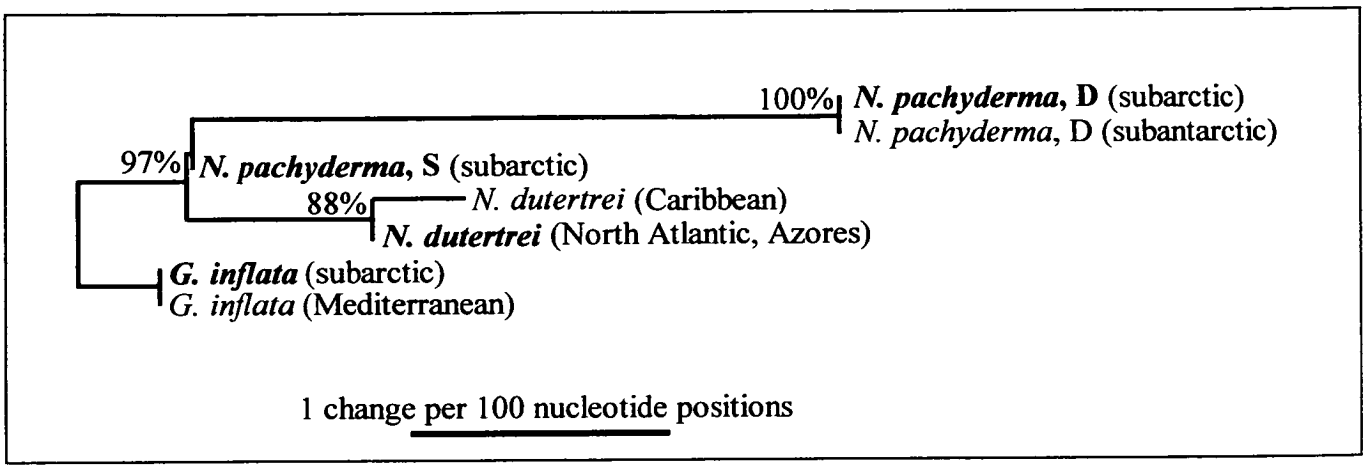
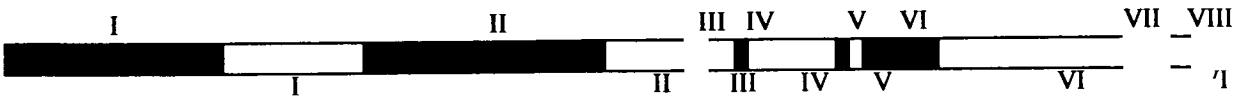


Fig. 7-1. Neighbour-joining phylogeny based on 284 unambiguously aligned sites, showing the relationships within the *Neogloboquadrina* cluster. *G. inflata* is used as an outgroup, as it diverged from a common ancestor prior to the formation of the *Neogloboquadrina* clade. Support has been assigned to the phylogeny using 1000 bootstrap replicates.

The 284 bp phylogeny (Fig. 7-1) indicates that *N. pachyderma* (S), *N. pachyderma* (D) and the *N. dutertrei* genotypes cluster within a single lineage with high bootstrap support (97 %). As already shown in the 505 bp phylogeny (Fig. 3-2), the association between the subarctic and subantarctic *N. pachyderma* (D) is supported in 100 % of the bootstrap replicates as these genotypes are identical throughout the entire ~1000 bp SSU rDNA fragment. The sinistral and dextral coiling forms of *N. pachyderma* are separated by an evolutionary distance of 2.5 % within the 284 bp phylogeny. Their association within this phylogeny was not resolved. The *N. dutertrei* genotypes from the Azores Current region of the North Atlantic and the Caribbean cluster together. Their association is supported in 88 % of the bootstrap replicates and they are separated by an evolutionary distance of 0.35 %. The Azores Current *N. dutertrei* is separated from *N. pachyderma* (D) and *N. pachyderma* (S) by evolutionary distances of 3.3 % and 0.7 % respectively. The

Caribbean *N. dutertrei* is separated from *N. pachyderma* (D) and *N. pachyderma* (S) by evolutionary distances of 3.6 % and 1.1 % respectively.

The phylogeny constructed using 284 bp unfortunately has limited resolution owing to the very few genetic differences in the alignable regions used for analysis (Fig. 7-2B). However, the variable regions of the partial SSU rDNA fragment are highly divergent (Figs. 2A and 2C). *Neogloboquadrina pachyderma* (D) is highly divergent from the other members of the clade, which severely reduces the number of alignable sites available for phylogenetic analysis. The high degree of genetic divergence is shown in Fig. 7-2C, where in all but one variable region, *N. pachyderma* (D) cannot be aligned, even against *N. pachyderma* (S).



A.

	Variable region sequence lengths (bp)						
Species	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**
N. pachyderma (D)	61	31	11	36	6	83	12
N. pachyderma (S)	52	35	13	38	4	62	11
N. dutertrei, Azores	38	35	12	32	4	55	10
N. dutertrei, Caribb.	41	35	12	32	4	55	10

B.

	Conserved regions used for phylogenetic analysis							
Species comparison	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**
N. pachyderma (D) / *N. pachyderma* (S)	2 bs	1 bs 1 del	0	0	0	0	2 bs	0
N. pachyderma (D) / *N. dutertrei*, Azores	2 bs	2 bs 1 del	0	0	0	1 bs	3 bs	0
N. pachyderma (S) / *N. dutertrei*, Azores	0	0	0	0	0	1 bs	1 bs	0
N. dutertrei, Azores / *N. dutertrei*, Caribb.	0	0	0	0	0	0	0	1 bs

C.

	Variable regions not used for phylogenetic analysis						
Species comparison	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**
N. pachyderma (D) / *N. pachyderma* (S)	U	U	U	U	U	U	2 bs 1 del
N. pachyderma (D) / *N. dutertrei*, Azores	U	U	U	U	U	U	1 bs 2 del
N. pachyderma (S) / *N. dutertrei*, Azores	U	5 bs	4bs 1 del	2bs 6 del	2 bs	U	1 bs 1 del
N. dutertrei, Azores / *N. dutertrei*, Caribb.	1 bs 3 del	0	0	0	0	2 bs	0

Fig. 7-2. Representation of the ~500 bp SSU rDNA fragment to illustrate the number of genetic differences between the *Neogloboquadrina* genotypes. The black represents the regions used for phylogenetic analysis, and the white represents the variable regions that were not used. **Fig. 7-2A** shows the sequence length variations within the variable regions of the region amplified. **Fig. 7-2B** shows the number of genetic differences that exist within the conserved regions used for phylogenetic analysis. **Fig. 7-2C** shows the number of genetic differences that exists within the variable regions not used for phylogenetic analysis. **'U'** denotes an unalignable region, **'bs'** denotes base substitution, and **'del'** denotes base deletion/insertion differences.

There is considerable sequence length variation between the genotypes, especially within the variable regions I and VI (Fig. 7-2A). Comparison of the

sequences (Fig. 7-2) shows consistency with the 284 bp molecular phylogeny (Fig. 7-1). There are significantly fewer genetic differences between *N. pachyderma* (S) and *N. dutertrei*, than between *N. pachyderma* (S) and *N. pachyderma* (D) or between *N. pachyderma* (D) and *N. dutertrei*, within both the conserved regions used for phylogenetic analysis (Fig. 7-2B) and also the variable regions (Fig. 7-2C). This observation is especially true for the variable regions, where 6 out of 7 variable regions are completely unalignable between *N. pachyderma* (S) and *N. pachyderma* (D) and between *N. pachyderma* (D) and *N. dutertrei*, indicating a significantly high level of divergence between the genotypes. Although considerable genetic differences exist between *N. pachyderma* (S) and *N. dutertrei* within the variable regions (Fig. 7-2C), there are much fewer than between *N. pachyderma* (S) and *N. pachyderma* (D) or between *N. pachyderma* (D) and *N. dutertrei*. In contrast, the two genotypes of *N. dutertrei* are very closely related, differing by only 1 base substitution within the conserved regions (Fig. 7-2B), and 3 base substitutions and 3 base deletion/insertions in the variable regions (Fig. 7-2C).

7.3. *Neogloboquadrina pachyderma* within the subarctic Atlantic

As described above, the sinistral and dextral coiling forms of *N. pachyderma* within the subarctic Atlantic are genetically distinct. Within a 284 bp phylogeny the sinistral and dextral coiling specimens are separated by an evolutionary distance of 2.5 % (Fig. 7-1). Within the variable regions not used for phylogenetic analysis the sequences are very different, with a number of unalignable regions (Fig. 7-2C).

The distribution of *N. pachyderma* genotypes collected within the subarctic Atlantic is shown in Fig. 7-3. The *N. pachyderma* (D) genotype was found to exist across the whole transect over a temperature range of >11°C- 2.5°C. The sinistral coiling specimen was obtained from station 4 (63°31N/39°43W). Surprisingly, even in the very cold water (2.5°C), across the polar front, the number of dextral specimens was at least equal to the number of sinistral specimens.

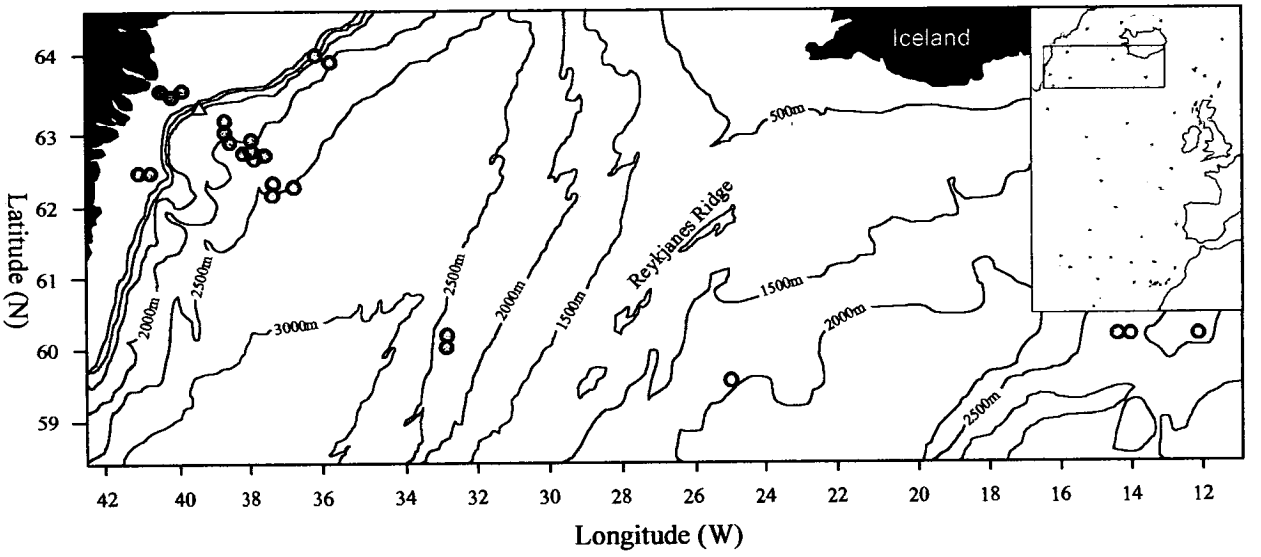


Fig. 7-3. Distribution of *N. pachyderma* genotypes within the subarctic Atlantic. The inset map indicates the enlarged area. Dextral specimens (24) are denoted in blue (O), and the single sinistral specimen is denoted in red (Δ).

7.3.1. Morphological variability of subarctic *N. pachyderma*

Neogloboquadrina pachyderma (dextral) displayed a range of morphological variability (Plate 7-1). Although the specimens shown appear to be morphologically similar to specimens of *T. quinqueloba* when viewing through a light microscope, they are clearly different when using SEM imaging (compare with Plate 4-3). The specimens shown are from bulk plankton collections and have between 4 and $4^1/_2$ chambers, with an apertural lip visible in the majority of specimens. A rarer 5

chambered specimen (not shown) was also sequenced and was found to be genetically identical to the other *N. pachyderma* (D) sequences. Not all specimens shown have reached their mature size and form.

Of the 24 specimens sequenced, the predominant cytoplasmic colour was bright red, accounting for over 50 % of the individuals sequenced (14), though orange (8) and yellow (2) were also observed. There was no correlation between colour and geographic distribution.

The *N. pachyderma* (S) specimen, from which the partial SSU rDNA sequence was obtained, displays a compact test morphology (Plate 7-1, Fig. 11). This image was recorded using a digital video camera mounted on a stereo microscope as described in Chapter 2 (section 2.2.4). If this specimen is compared to the dextral coiling specimens (Plate 7-1, Figs. 1-10), it is apparent that they differ in gross test morphology due to the more compact test of the sinistral specimen (except perhaps for the dextral specimen shown in Plate 7-1, Fig. 2). For reference, two other sinistral specimens obtained from the bulk plankton samples are shown.

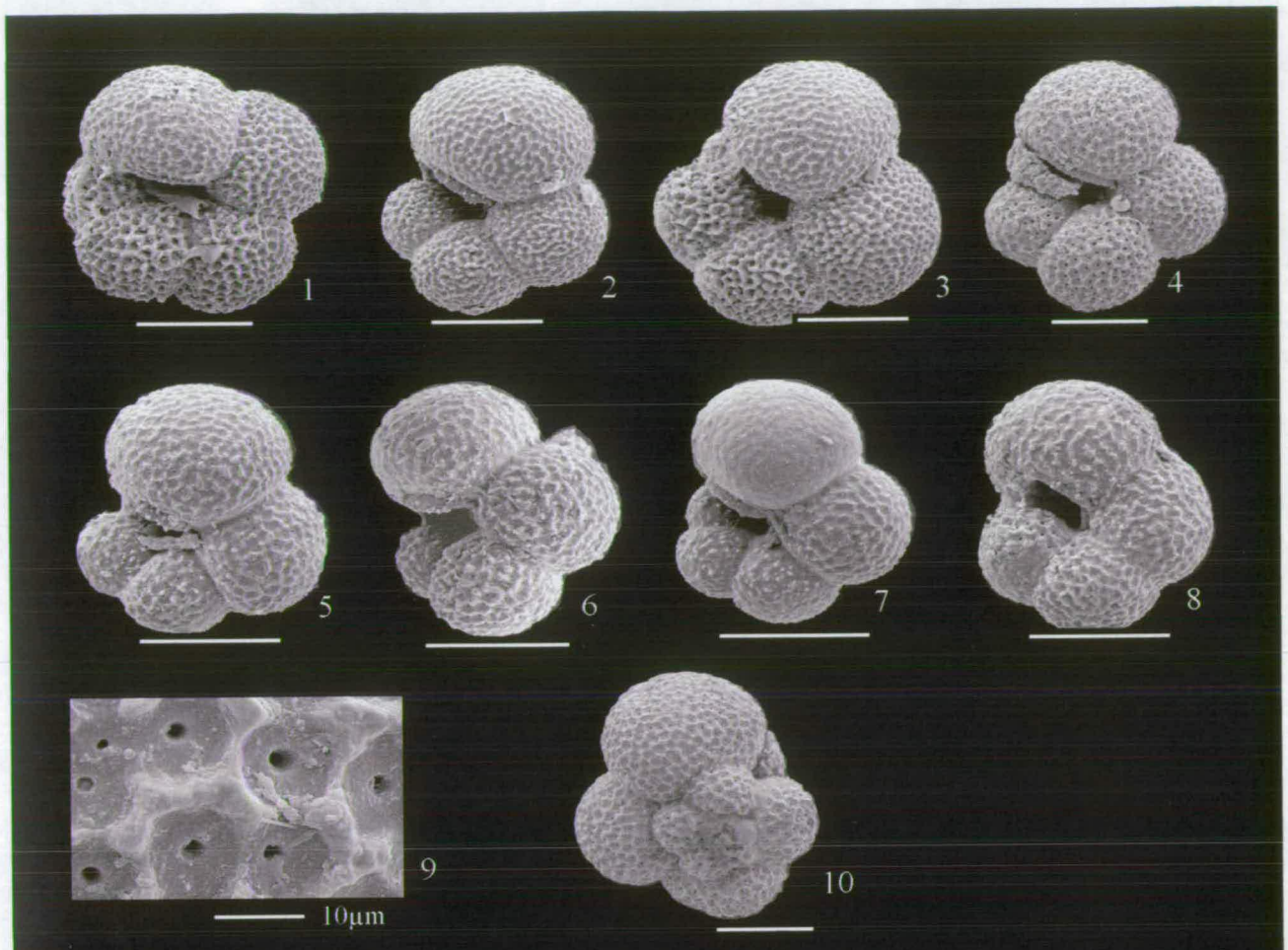
Plate 7-1 description: *Neogloboquadrina pachyderma* (Ehrenberg) from the subarctic Atlantic. Dextral (Figs. 1-10) and sinistral (Figs. 11-15) specimens are shown. Unless otherwise indicated the scale bars represent 100 μm.

Dextral specimens:

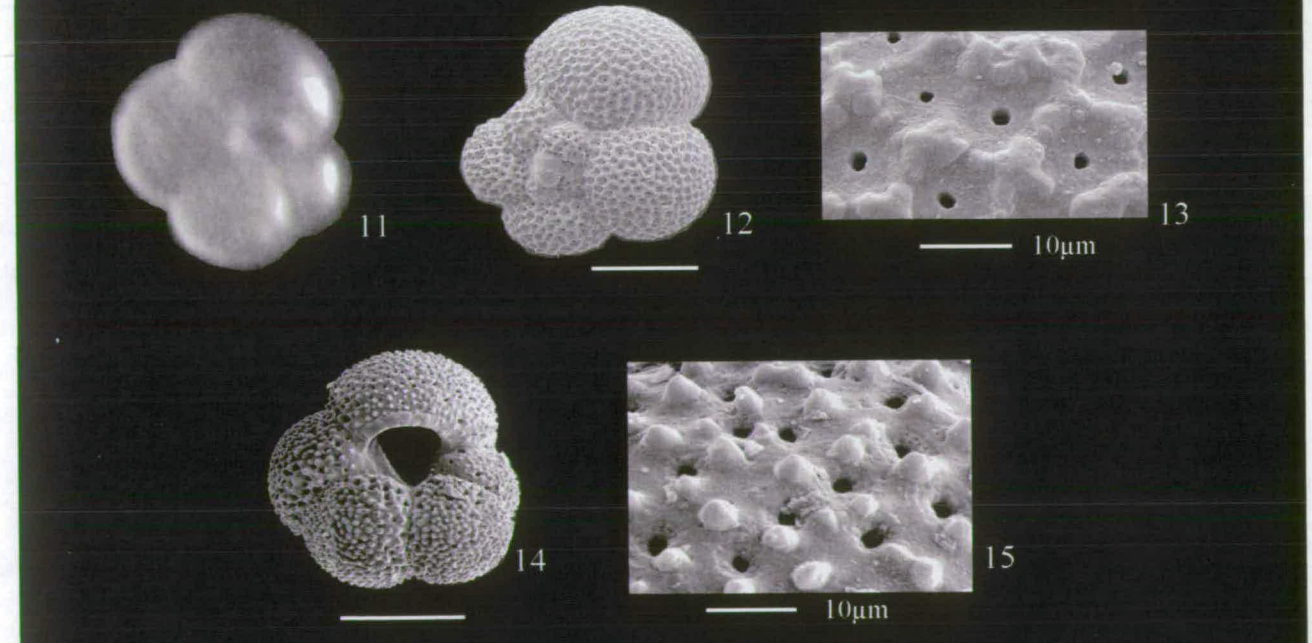
Fig. 1: 4 ½ chamber specimen, umbilical side, possibly apertural lip visible on final chamber. **Fig. 2**: 4 chamber specimen, umbilical side, note apertural lip on last chamber. **Fig. 3:** 4 ¼ chamber specimen, umbilical side, note heavy calcification on the 2nd and 3rd last chambers. **Figs. 4-6:** 4 chamber specimen, umbilical side, apertural lip visible. **Fig. 7:** 4 ½ chamber specimen, umbilical side, apertural lip. Note smooth, newly formed final chamber. **Fig. 8:** 4 chamber specimen, umbilical side, upper apertural lip thickened, chambers appear to be more "joined" than the other specimens. **Fig. 9:** same specimen as in Fig. 4, enlargement of final chamber. Note the large pores, and the inter-pore ridges. **Fig. 10:** trochospiral side of specimen in Fig. 4.

Sinistral specimens:

Fig. 11: digital video image of 4 chamber specimen from which partial SSU rDNA sequence was obtained, umbilical side. **Fig. 12:** trochospiral side, similar morphology to the dextral coiling specimen shown in Fig 10. **Fig. 13:** same specimen as in Fig. 12, enlargement of final chamber. Note the large pores and the developing inter-pore ridges. **Fig. 14:** 5 chamber specimen, umbilical side, thickened apertural rim, significant calcification on chambers. **Fig. 15:** same specimen as in Fig. 14, enlargement of final chamber. Note very large pore sizes, and the calcite nodules lying between them.



Neogloboquadrina pachyderma (D) morphotypes



Neogloboquadrina pachyderma (S) morphotypes

Plate 7-1

7.4. *Globigerinita uvula*

Four specimens were obtained during the subarctic transect from the area south of Iceland and the east coast of Greenland (Fig. 7-4). Two further specimens were obtained from the transitional waters north of Scotland (59°45.7'N/05°46.9'W). All six specimens were identical throughout the partial SSU rDNA fragment.

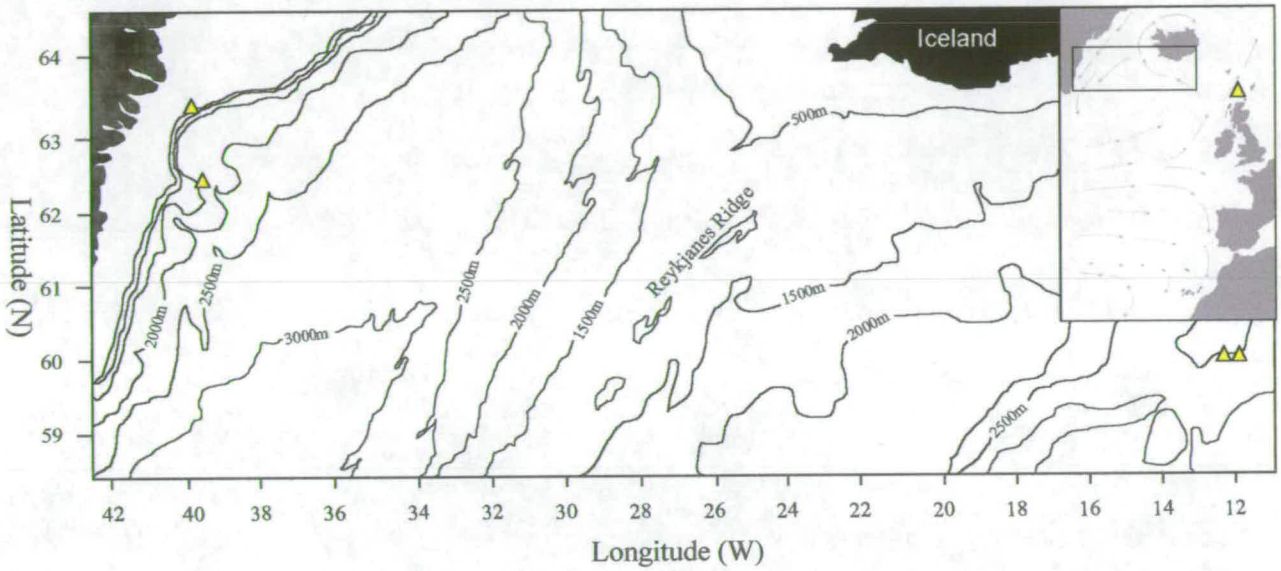


Fig. 7-4. Location of *G. uvula* (△) genotypes obtained from the subarctic Atlantic. The inset map indicates the enlarged area, and the triangle denotes the approximate locality of the 2 further specimens.

This microperforate species had a small, generally smooth, trochospiral test. Pustules were observed on the larger chambers (Plate 7-2). Each live specimen had a characteristic bright yellow cytoplasmic colour.

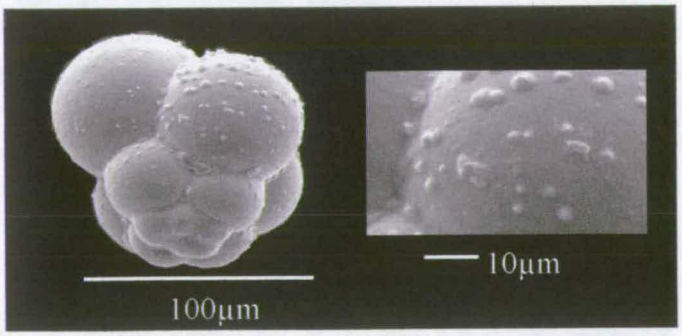


Plate 7-2. Morphology of *Globigerinita uvula*. The magnified view of the test illustrates its smooth texture. Note the pustules present on its surface.

7.5. Discussion

The *N. pachyderma* (D) and *G. uvula* genotypes fall separately within the benthic and non-spinose planktic region of the foraminiferal molecular phylogeny (see Chapter 3), indicating that they evolved from different benthic ancestors.

The molecular phylogeny based on 284 bp (Fig. 7-1) shows that subarctic Atlantic sinistral and dextral *N. pachyderma* are genetically distinct. Following morphological analysis of the *neogloboquadrinids* it has generally been the accepted view that *N. pachyderma* (D) was an intergrade between the *N. pachyderma* (S) and *N. dutertrei* (Bé and Tolderlund, 1971). However, the molecular data (Fig. 7-1 and Fig. 7-2) has shown this interpretation to be incorrect, since *N. pachyderma* (S) has a closer evolutionary relationship to *N. dutertrei* than to *N. pachyderma* (D). This is in agreement with the findings of Darling *et al.* (submitted) for Antarctic *N. pachyderma.* Therefore, coiling direction in subarctic *Neogloboquadrina pachyderma* is associated with genetically distinct types, and is not a developmental trait determined by temperature. Further molecular analysis of *N. pachyderma* (S) within the high latitude North Atlantic would permit an investigation of whether all sinistral specimens were genetically homogeneous, or whether there are genetically distinct types as found within the polar/subpolar regions of the South Atlantic (Darling *et al.*, submitted).

Within the subarctic Atlantic, a single *N. pachyderma* (D) genotype spanned the whole transect, being found across the entire temperature range from 11°C down to 2.5°C. This implies that temperature does not play a major role in the distribution of this genotype in the subarctic North Atlantic. However, it is possible that the *N. pachyderma* (D) specimens found near the East Greenland margin are not actually

reproducing there, but instead are expatriates from the Irminger Current. It must be noted that the extensive distribution of *N. pachyderma* (D) concurs with Cifelli (1971) but is in contrast to the distributions found during other subpolar/polar plankton collections. In these cases a shift from dextral to sinistral coiling was observed at approximately 7-8°C (Bé, 1960; Bé and Hamlin, 1967). The interpretations of the fossil record are rather confused with respect to *N. pachyderma* coiling provinces in the subarctic Atlantic. Ericson (1959) estimated that the shift between dextral and sinistral coiling in this region corresponded to the 7.2°C April isotherm. However, the paper is somewhat contradicting since a core top from within the region of influence of the Irminger Current, which is supposedly in the sinistral province, has a 70 % dominance of dextral specimens (Ericson, 1959). This supports the suggestion that dextral specimens near the East Greenland margin may be expatriates from the Irminger Current.

The lack of sinistral specimens in the cold waters of the east Greenland margin (this study) and the Labrador Current (Cifelli, 1971) may be related to salinity. In these regions the seawater is less saline (see Fig. 4-15) due to the input of freshwater from rivers and also glacial melt. Although *N. pachyderma* (S) is a polar water inhabitant it may not tolerate very low salinity water, which could explain its low numbers in these areas. Therefore, due to the influence of low salinity water near the Greenland margin, perhaps the shift in *N. pachyderma* coiling direction at 7-8°C is more applicable to the open ocean. However if this were the case it would be an unusual adaptation. Indeed, this would be in contrast to the results of the culture experiments on planktics by Bijma *et al* (1990a) which showed that salinity tolerance, in the species studied, was in excess of that found in the oceans. In

addition, if the *N. pachyderma* (D) specimens at the east Greenland margin are not expatriates from the Irminger Current but reproduce there, they are clearly tolerant to a wide salinity range. Investigation of the habitat tolerance of *N. pachyderma* (D and S) using culture experiments would aid the interpretation of these data.

Neogloboquadrina pachyderma (D) displayed a range of morphological variability (Plate 7-1), although not as extensive as subarctic *T. quinqueloba*. In contrast to *T. quinqueloba* however, all the *N. pachyderma* (D) genotypes were genetically identical throughout the whole transect. If *N. pachyderma* (D) proves to be genetically homogeneous, then any morphological variation outwith the observed morphological range may provide an indicator of environmentally induced phenotypic variation. The *N. pachyderma* morphotypes described in Plate 7-1 are similar to the *N. incompta* morphotypes described by Cifelli (1961), indicating that *N. incompta* is synonymous with *N. pachyderma* (D). Further, the *N. dutertrei* genotype from the subtropical N. Atlantic was found to be distinct from the tropical *N. dutertrei* genotype. Additional work will be required to determine the distribution of this genotype within the North Atlantic, and its precise evolutionary relationship to the tropical *N. dutertrei*. The genetic divergence between subtropical and tropical *N. dutertrei* is supported by the conclusions of Hilbrecht (1997) which indicated that in the North Atlantic the subtropical population of *N. dutertrei* appears to be geographically isolated from the tropical population. These data may prove valuable in resolving the *N. pachyderma* (D) / *N. incompta* Cifelli / *N. pachyderma* – *N. dutertrei* intergrade group (Hilbrecht, 1997). Although an identical *N. pachyderma* (D) genotype has been found in the subpolar Antarctic waters (Darling *et al.*,

submitted), it remains to be determined whether other *N. pachyderma* (D) genotypes of similar morphology will be found.

The discovery of identical *N. pachyderma* (D) genotypes within the subarctic and subantarctic has considerable implications for gene flow between the poles. As this species has a predominantly bipolar (antitropical) distribution, it is apparent that there has been recent gene flow between these two subpolar populations across the tropical waters of the Atlantic Ocean. In conjunction with other bipolar data from *G. bulloides* and *T. quinqueloba*, this will be considered in detail in Chapter 8.

Chapter 8: Discussion Index

8.1. Genotypic variation in North Atlantic planktic foraminifera: cryptic speciation ?

8.2. Gene flow

8.2.1. Bipolarity/antitropicality

8.2.1.1. Phylogenetic relationship between bipolar planktic foraminiferal morphospecies

8.2.1.2. Mechanism of genetic exchange between subpolar populations

8.2.2. Genetic exchange between North Atlantic and Pacific transitional/subtropical planktic foraminifer populations

8.2.2.1. The degree of evolutionary relatedness between North Atlantic and Pacific Ocean genotypes

8.2.2.2. What are the possible routes into and out of the North Atlantic?

8.2.2.3. Species tolerance to movement between the Pacific and Atlantic Oceans

8.2.2.4. Habitat tolerance

8.2.2.5. Potential scenarios for global isolation and speciation

8.2.2.6. Conclusions

8.3. Implications and further work

8.3.1. Taxonomy and phylogenetic implications

8.3.2. Application of molecular data to palaeoceanography

8.1. Genotypic variation in North Atlantic planktic foraminifera: cryptic speciation ?

Molecular phylogenetic advances have permitted detailed investigation of morphologically defined marine species. It is apparent that biodiversity within the oceans is much higher than ever imagined as many marine morphospecies have been found to represent more than a single species (see review by Knowlton, 1993). The terms sibling species, sister species and cryptic species have been used to describe this phenomenon, and for the purposes of discussion they are considered synonymous. The term "cryptic speciation" refers to biological speciation that has not resulted in a discernible morphological change (Huber *et al.*, 1997); hence cryptic species cannot be distinguished by traditional palaeontological methods in the fossil record. At its simplest, biological speciation can be thought of in terms of the establishment and maintenance of reproductively isolated populations, as first discussed by Dobzhansky (1937).

Knowlton (1993) observed that sibling species were ubiquitous within the marine environment, being common in marine invertebrates, marine vertebrates and plants. Indeed, on the basis of molecular phylogenetic evidence, it is now recognised that present-day planktic foraminiferal diversity is much higher than fossil record interpretations have previously suggested (Darling *et al.*, 1997, 1999). A number of morphologically defined planktic foraminiferal species have been found to consist of genotype clusters (Darling *et al.*, 1999), and it has been proposed that some of these genotypes may also represent cryptic species (Huber *et al.*, 1997; Darling *et al.*, 1997, 1999; de Vargas *et al.*, 1999).

Cryptic speciation has also been reported from the fossil record and has been suggested as the reason behind the shift in reproductive depth of the *Globorotalia* (*Fohsella*) lineage, as identified by Norris *et al.* (1996) using isotopic data. Comparison of the 505 bp molecular phylogeny with interpretations of the fossil record show that, in a number of cases, the first appearance of some planktic foraminiferal morphospecies is earlier than the fossil record indicates (Chapter 3). Unfortunately, large variations in evolution rate exist between, and within, monophyletic groups (Fig. 3-2) which makes calibration of the molecular phylogeny problematic. If a constant rate of molecular evolution is assumed within individual clusters of the molecular phylogeny and a datum level imposed as calibration, the molecular phylogeny always indicates divergences prior to those indicated by the fossil record. For example, within the *Globigerinella* lineage there is a considerable difference between the first recorded appearance of *G. calida* within the fossil record (~ 4 Ma; Kennett and Srinivasan, 1983) and its evolutionary divergence datum as suggested by the molecular phylogeny (10-11 Ma; Chapter 3, section 3.4.2). This is likely to be due to the molecular divergence at speciation preceding morphological change, and therefore remaining cryptic to micropalaeontological investigation (see Chapter 3).

8.1.1. What level of genetic distinction is required to propose that genotypes are cryptic species?

Ideally, it would be preferable to have more evidence than genetic distance of the SSU rRNA gene alone. Indeed, the use of other biological data and/or alternative genes would be highly desirable to corroborate the molecular phylogenetic data

inferred from the SSU rRNA gene. However, it can be supposed that genetic distance might be indicative of speciation. There is not, unfortunately, a specific level of genetic distance that can be taken as indicating a species-level distinction, due to the considerable variation in evolution rate between morphospecies. In addition, although the level of genetic distance observed between two genotypes will reflect time, a small genetic distance may reflect slow evolution rates and/or a recent speciation event. Indeed, it has been recognised by Knowlton (1993) that sibling species may actually show very little genetic differentiation.

The level of genetic distance observed between genotypes is specific to the molecular phylogeny, since genetic distances vary according to the number of unambiguously aligned sites used to reconstruct the phylogeny. Therefore, comparison of genetic distance between molecular phylogenies using different numbers of sites is misleading.

Recently, using morphological, biological or chemical corroborative evidence, the genetic variation between the *G. siphonifera* Type I and Type II genotypes (Huber *et al.*, 1997; Darling *et al.*, 1997), between the *Orbulina universa* genotypes (de Vargas *et al.*, 1999), and between the *G. ruber* pink and white genotypes (Darling *et al.*, 1999) has been taken as representing species level differences. Within the 505 bp molecular phylogeny presented in Chapter 3, an evolutionary distance of 8.6 % separates two of the cryptic *O. universa* species, an evolutionary distance of 3.5 % separates *G. siphonifera* Types I and II, and an evolutionary distance of 5-10 % separates *G. ruber* pink from the white genotypes. It is highly likely that a smaller evolutionary distance between some genotypes within a

morphospecies also reflects species level distinction, as the smaller genetic distance will simply reflect a shorter period of time since the speciation event.

8.1.2. Molecular evidence for cryptic speciation from the North Atlantic

Within the North Atlantic, genotypic variation has been identified within the following planktic foraminiferal morphospecies.

Globigerina bulloides – 3 genotypes

Turborotalita quinqueloba – 2 genotypes

Globigerinella siphonifera – 3 genotypes

Globigerinella calida – 2 genotypes

Globigerinoides ruber – 3 genotypes

Neogloboquadrina pachyderma – 2 genotypes

In a conservative molecular phylogeny, based on 505 unambiguously aligned sites (Fig. 3-2), the level of genetic divergence within a morphospecies ranges considerably (see Chapters 4, 5, 6, and 7). In the 505 bp molecular phylogeny the evolutionary distance between genotypes within these genotype clusters ranges from less than 1 % to over 10 % (Table 1).

Morphospecies	Genotype comparison	Evolutionary distance (%)
G. bulloides	Type Ib vs. Type II (a+b)	8.2-9.2
G. bulloides	Type IIa vs. Type IIb	1.0
G. siphonifera	Type I vs. Type II (a+b)	3.3-3.7
G. siphonifera	Type IIa vs. Type IIb	1.2
T. quinqueloba	Type IIa vs. Type IIb	0.6
G. ruber	Pink vs. white Type I	5.2
G. ruber	Pink vs. white Type II	10.2
G. ruber	white Type I vs. white Type II	9.4

Table 1. Genetic divergences within North Atlantic morphospecies, based on the conservative 505 bp molecular phylogeny (Chapter 3).

The most divergent genotypes in the 505 bp molecular phylogeny are found within *G. ruber* and *G. bulloides* (Table 1). The distances reflect both the rate of molecular evolution and also the period of time since divergences occurred. However, the conservative 505 bp molecular phylogeny cannot resolve the actual genetic divergence between genotypes within clusters, as much genetic divergence is found in the variable regions which cannot be unambiguously aligned between morphospecies. Resolution can be improved by reconstructing a molecular phylogeny using only genotypes from a single morphospecies (Table 2).

Morphospecies	Genotype comparison	Phylogeny (bp)	Evolutionary distance (%)
G. bulloides	Type IIa vs. Type IIb	935	1.2
G. siphonifera	Type I vs. Type II (a+b)	767	6.9-7.1
G. siphonifera	Type IIa vs. Type IIb	767	2.0
T. quinqueloba	Type IIa vs. Type IIb	762	1.6

Table 2. Genetic divergences within North Atlantic morphospecies, based on the molecular phylogenies reconstructed using only genotypes from a single morphospecies.

Whilst the evolutionary distances between most of the genotypes still remained relatively low, the distance between *G. siphonifera* Type I and Type II (a and b) increased considerably (Table 2). However, even the within-morphospecies molecular phylogeny can mask the true level of sequence divergence that exists between some genotypes. Where little divergence exists in the conserved region, close comparison of the variable regions of the SSU rDNA fragments permits a detailed examination of the genetic variation that exists between genotypes, although any statistical support is inevitably lost. At this level, the North Atlantic genotypes within a morphospecies are often more highly divergent from one another than phylogenetic analysis shows (see Chapters 4, 5, 6, 7).

It is apparent from the molecular phylogenetic data, that some of the genotypes identified within North Atlantic morphospecies must be cryptic species. However, since there is not a definitive evolutionary distance that characterises species level distinction, caution must be taken unless supported by additional biological or chemical evidence. The genetic distances between the genotypes ranges considerably, but some are more divergent than other described cryptic species (e.g. de Vargas *et al.*, 1999). In addition to the species level distinction already proposed for *G. siphonifera* Types I and II (Huber *et al.*, 1997; Darling *et al.*, 1997) and *G. ruber* pink and white (Darling *et al.*, 1999), the genetic distances are of a sufficient magnitude to propose that the *G. bulloides* Type Ib genotype is a distinct species from the *G. bulloides* Type II genotypes and that the *G. ruber* white Type I and Type II genotypes are distinct species. Comparison of the foraminiferal molecular phylogeny with interpretations of the fossil record, by Darling *et al.* (1999), suggested that the two *G. ruber* white lineages may have diverged as long ago as ~22 Ma.

The closer evolutionary relationships observed within the *G. bulloides* Type II cluster, the *T. quinqueloba* Type II cluster and the *G. siphonifera* Type II cluster is more problematic. It is possible that these genotypes are also indicative of cryptic speciation and simply reflect a more recent divergence. Taking their molecular evolution rates into consideration, it is possible that the genotypes from these clusters diverged during the Quaternary. However, it is too early to make such conclusions from molecular data alone.

The proposal that some of the genotypes may be cryptic species is supported by the fact that the genotypes are discrete. No obvious hybrid genotypes have been

identified, indicating that the genotypes are reproductively isolated as they are often found together in the water column. This is of key importance, since genetic divergence must be accompanied by reproductive isolation for speciation to occur (Coyne and Orr, 1998).

8.1.3. Ecological evidence

It is evident from the molecular phylogenetic data that genetic divergences within the SSU rRNA gene of North Atlantic planktic foraminiferal morphospecies are quite common. The next question to address is whether the genotypes within a morphospecies have different distributions in the North Atlantic, as this may reflect habitat preferences and evolutionary adaptation. Such adaptation, if found, has fundamental significance for palaeoceanographic investigations hence the distribution of genotypes was investigated in an attempt to address this issue.

8.1.3.1. Spatial distribution of genotypes in the North Atlantic

(A) *Globigerina bulloides* (see Chapter 4). In the subarctic Atlantic, two *G. bulloides* genotypes (Types IIa and Type IIb) were obtained. Along the subarctic transect a distributional change was found, from a population composed predominantly of *G. bulloides* Type IIb genotypes (east of the Reykjanes Ridge), to a population composed entirely of *G. bulloides* Type IIa genotypes (west of the Reykjanes Ridge). A transition zone between the two genotype populations approximates to the Reykjanes Ridge (Fig. 4-6). However, in the transitional-subtropical North Atlantic, three *G. bulloides* genotypes (Type Ib, Type IIa and Type IIb) were found to co-exist within the NAC (Fig. 4-7). The southern-most occurrence

of *G. bulloides* Type IIa was 43°N, whilst *G. bulloides* Type IIb was found in the subtropical waters of the AC at 35°N. In addition to being found in the NAC, *G. bulloides* Type Ib was also obtained from the subtropical waters of the Canary Islands (Fig. 4-7). The Type Ib genotype is also found within the Mediterranean Sea (de Vargas *et al.*, 1997). However, the Type Ib genotype was not found to be present in the subarctic region.

(B) *Turborotalita quinqueloba* (see Chapter 4). In the subarctic region of the Atlantic, two *T. quinqueloba* genotypes (Type IIa and IIb) were obtained. In the first section of the subarctic transect (east of the Reykjanes Ridge) both genotypes co-existed within the water column (Fig. 4-11). In the second section of the subarctic transect (west of the Reykjanes Ridge) only *T. quinqueloba* Type IIa was found. However, due to low sampling numbers in this region, further investigation is required to confirm that this is a homogeneous population.

(C) *Neogloboquadrina pachyderma* (see Chapter 7). In the subarctic region of the Atlantic, two *N. pachyderma* genotypes were obtained. These genotypes corresponded to the sinistral (S) and dextral (D) coiling variety of this morphospecies. A single *N. pachyderma* (D) genotype spanned the subarctic transect, and the *N. pachyderma* (S) genotype was obtained from near the east Greenland margin (Fig. 7-3). The numbers of sinistral specimens were very low during the subarctic collection, even in the very cold waters associated with the East Greenland Current (EGC).

(D) *Globigerinella* sp. (see Chapter 5). No *Globigerinella* sp. were found in the subarctic region of the Atlantic. In the transitional-subtropical region of the North Atlantic, three genotypes of *G. siphonifera* (Type I, Type IIa and Type IIb) and two genotypes of *G. calida* were obtained. The *G. siphonifera* Type I genotype was rare, being found only in the waters just north of the Azores frontal zone and the subtropical waters of the Canary Islands (Fig. 5-6). The *G. siphonifera* Type IIa and Type IIb genotypes were found extensively across the transitional-subtropical water masses of the region. The distribution of the *G. siphonifera* Type IIa and Type IIb genotypes was similar, and they were found to co-exist in the water column (Fig. 5-6). One *G. calida* genotype was found in the NAC and the NATW, and the variant genotype was found in the AC. The numbers of *G. calida* obtained from the subtropical waters were very low (Fig. 5-7).

(E) *Globigerina falconensis* (see Chapter 4). In contrast to the genetic variability found within the morphospecies described above, *G. falconensis* was represented by only a single genotype in the transitional-subtropical region of the North Atlantic. Its range stretched from the NAC to the AC (Fig. 4-14), but does not extend into the subarctic region.

The genotype distribution pattern is quite complex and morphospecies specific. It is evident that genotypes within a morphospecies often co-exist within the water column. This supports the molecular evidence for cryptic speciation since the maintenance of distinct co-existing genotypes must require reproductive isolation, which is the basis for biological speciation. Further, the genotypes within a

morphospecies also often differ in their spatial distribution, which indicates that there is an element of habitat preference/tolerance involved.

These distribution data will require further confirmation due to the low sampling density in some cases, but a number of points regarding the distribution of specific genotypes stand out. The *G. bulloides* Type IIa and Type IIb genotypes display an extensive geographic range, especially the Type IIb genotype since it was found from the subarctic to the subtropical AC. It is quite possible that these "cool" water genotypes are actually generalist species, capable of tolerating a wide range of oceanic conditions. However, it is evident that *G. bulloides* Type Ib does not inhabit the subarctic waters of the North Atlantic, suggesting that it is cold intolerant. It is possible that while the Type Ib genotype can tolerate the cool NAC, it is perhaps more of a "warmer" water specialist. Similarly, *G. siphonifera* Type I was only found in low numbers in the warmer waters of the transitional-subtropical region, suggesting that it prefers warmer waters and/or different productivity conditions. On the other hand, the *G. siphonifera* Type II genotypes were found extensively across the transitional-subtropical region of the North Atlantic. In Chapter 5 (section 5.5.1) it was proposed that *G. siphonifera* Type I may be specialised to warm oligotrophic waters, whereas the *G. siphonifera* Type II genotypes may be more generalist, capable of living in transitional/subtropical waters with variable productivity levels.

8.1.3.2. Depth distribution

The depth distribution of planktic foraminiferal morphospecies has been examined using multi-net plankton tows (e.g. Fairbanks *et al.*, 1980, 1982; Ottens, 1991, 1992; Schiebel *et al.*, 1997). The data collected in this study has not provided

any information regarding the depth habitat of genotypes within the water column. However, multi-net samples (0–700 m) were obtained from the Azores current system during the Poseidon cruise (Schiebel, 1999). The planktic foraminiferal assemblages obtained by pump samples from the surface waters (this study) were comparable to the multi-net assemblages. The only difference between the two collection methods was that the diversity of planktic foraminiferal morphospecies obtained from the pump samples was higher than in multi-net samples. During much of this cruise there were high winds, which made the sea rough and increased the vertical mixing within the water column. The high level of vertical mixing was reflected in the presence of mature *G. truncatulinoides* in the surface waters, since mature specimens are thought to be deep-dwellers (Hemleben *et al.*, 1989). Therefore the surface water is an adequate representation of the upper water column.

8.1.3.3. Factors affecting genotype distribution

(1) Historical chance effects

Without further resampling, the possibility that the distribution patterns observed are as a result of historical chance effects cannot be ruled out. However, the genotypes of *G. bulloides*, *T. quinqueloba*, *G. siphonifera* and *N. pachyderma* do have different distributions, suggesting that this seems unlikely. In addition, specific genotypes have been identified in different regions of the globe (Darling *et al.*, 1999), and de Vargas et al. (1999) found that the *O. universa* genotype distribution within the North and South Atlantic was strongly associated with chlorophyll levels, which is suggestive of adaptation. Therefore, the genotype distribution patterns are

more likely to be the result of a combination of variations in ocean parameters such as temperature and primary productivity, rather than historical chance effects.

(2) Seasonality

Seasonality is reflected in changes in the ocean parameters, such as temperature, salinity and primary productivity. As the seasons change the planktic foraminiferal morphospecies assemblage composition and relative abundance is known to vary (e.g. Deuser *et al.*, 1981; Deuser and Ross, 1989), reflecting the morphospecies' preference/tolerance for particular ocean characteristics. Comparison of observations from this study with other published data indicates some significant contrasts in species abundance during different times of the year. For example, no *T. quinqueloba* specimens were obtained from the NAC or NATW in January (this study) whereas they were abundant in this region during April and August (Ottens, 1991, 1992). Similarly, the relatively low numbers of *G. bulloides* found in the NAC in January (this study) is in contrast to their high abundance during April and August (Ottens, 1991, 1992). Further, the abundance of *G. falconensis* within the AC system is much higher in January than in August (Schiebel, 1999). If genotypes within a morphospecies do have specific habitat preferences/tolerances, it is highly likely that they will vary in relative composition and abundance during the course of the seasons.

(3) Temperature

The effects of temperature on morphospecies abundance are well recorded, and culture experiments have shown that the temperature tolerance of planktic

foraminiferal morphospecies in culture is comparable to their distributional range found in the ocean (Bijma *et al.*, 1990a). It is possible that genotypes within a morphospecies are affected in the same way. For example, the distribution of *G. bulloides* Type IIa and Type IIb in the subarctic region is indicative of different habitat tolerances. The *G. bulloides* Type IIa genotype is predominantly associated with the colder waters (as low as 2.5°C) of the subarctic region, towards the east Greenland margin. However, the *G. bulloides* Type IIb genotype was not found west of the Reykjanes Ridge, suggesting that it may not tolerate these colder waters.

The genotype distribution data suggests that there are "generalist" genotypes (e.g. *G. bulloides* Types IIa and IIb and *G. siphonifera* Types IIa and IIb) which can tolerate a wide range of SSTs, and more "specialist" genotypes (e.g. *G. bulloides* Type Ib and *G. siphonifera* Type I) which tolerate a narrower range of SSTs. However, as the data suggests, there may be a further underlying complexity to the genotype distributions.

(4) Salinity

The laboratory study of planktic foraminiferal morphospecies tolerances, by Bijma *et al.* (1990a), showed that salinity does not appear to be a limiting factor for morphospecies distribution in the oceans, since the morphospecies examined could all tolerate salinities far in extreme of those found in the ocean. However, the low numbers of *N. pachyderma* (S) obtained from the cold waters (2.5°C) near the east Greenland margin (this study) and the Labrador Current (Cifelli, 1971) suggest that perhaps salinity does have an effect on the sinistral type. As discussed in Chapter 7 (section 7.5) it is possible that the sinistral type does not tolerate very low salinity

water, which is characteristic of the east Greenland margin due to the low salinity of the prevailing EGC (Fig. 4-15). This could account for the low abundance of *N. pachyderma* (S), even in cold waters where it would be expected to be dominant.

(5) Primary productivity

It is known that ocean primary productivity (as represented by the chlorophyll concentration) is important with regard to planktic foraminiferal species abundance (Sautter and Thunell, 1991a; Deuser and Ross, 1989; Kroon and Ganssen, 1989; Ortiz *et al.*, 1995). During the subarctic collection (August/September) the primary productivity level is at its highest, whereas during the transitional/subtropical collection (January) the primary productivity level is at its lowest (Fig. 4-16). However, chlorophyll concentration in the oceans is patchy due to mesoscale eddies (Washburn *et al.*, 1998), causing it to vary both spatially and temporally.

Productivity is possibly reflected in the abundance of *G. bulloides* within the water column, since *G. bulloides* is known to prefer high productivity conditions (e.g. Sautter and Thunell, 1991a). This would account for the very high numbers of *G. bulloides* specimens obtained from the subarctic Atlantic during August/September, and the relatively few *G. bulloides* specimens obtained from the transitional North Atlantic during January. It is likely that productivity is perhaps one of the most important ocean parameters controlling genotype distribution. Indeed, de Vargas *et al.* (1999) identified a link between *O. universa* genotype distribution and ocean productivity. Unfortunately, due to its patchiness, linking chlorophyll levels to genotype distribution will only be achieved with *in situ* chlorophyll measurements.

The main problem in trying to relate the genotype distribution data to the ocean environment is that it only represents a "snapshot" in time. The spatial and temporal changes in oceanic parameters such as temperature, salinity and productivity make it difficult to form a specific link between genotype distribution and the ocean environment. The only way in which this could be achieved would be by long-term sampling in specific study areas, combined with detailed measurements of ocean parameters such as temperature, salinity and productivity. However, in many cases the genotypes within a morphospecies do have different spatial distributions within the North Atlantic, and the data suggests the presence of "generalist" and "specialist" genotypes. This points to habitat adaptation and supports the proposal that some of the genotypes within a morphospecies are cryptic species.

8.1.4. Genotype/morphotype relationships and other potential proxies

As described in Chapter 2, the specimens collected for genetic analysis are destroyed in the process, and the attempt to record digital video images of these specimens had limited success. Therefore, an investigation of genotype/morphotype relationships was attempted using bulk collected plankton samples. However, the bulk plankton samples yielded many juvenile specimens, with very few mature individuals. Distinguishing the juvenile specimens proved difficult as the morphology of specimens changes considerably during ontogeny. In addition, palaeoceanographic investigations utilise mature specimens from the marine sediments. Therefore, any genotype/morphotype investigation must relate to mature specimens to be of use to palaeoceanographers. If a significant number of mature

specimens had been obtained from bulk plankton samples in this study, a morphometric analysis could have been carried out.

Some morphospecies are represented by a single genotype across an ocean transect (e.g. *G. falconensis*). The range of morphology identified across this transect is therefore representative of this genotype. However, there are serious problems in using bulk samples to represent specific genotypes when confounded by the presence of multiple genotype populations. For example, three *G. bulloides* genotypes co-exist within the NAC (Fig. 4-7), three *G. siphonifera* genotypes (Fig. 5-6) and three *G. ruber* genotypes (Chapter 6) co-exist within the waters north of the Canary Islands and two *T. quinqueloba* genotypes co-exist within the subarctic Atlantic (Fig. 4-11). Unless homogeneous genotype populations are found, the linking of genotype to morphotype will be always be problematic using the bulk sampling approach.

There are, however, some genotypes that can be discriminated solely from their morphological characteristics. The first is *G. ruber* pink from *G. ruber* white (Chapter 6). To date *G. ruber* pink has only been represented by a single genotype. The second is *N. pachyderma* (D) from *N. pachyderma* (S) (Chapter 7), and to date, only a single genotype of *N. pachyderma* (D) has been found in both the subarctic and subantarctic regions (Darling *et al.*, submitted).

By definition, cryptic species are difficult to distinguish morphologically (Huber *et al.*, 1997) and even already defined morphospecies can be very similar in gross test morphology. For example, the morphospecies *G. bulloides* and *G. falconensis* have maintained a very similar gross test morphology, although they have been divergent for at least 18 Ma (Kennett and Srinivasan, 1983). It is quite possible for genotypes within a morphospecies to display little morphological

differences, since molecular evolution may precede morphological evolution. In addition, it is possible that morphological evolution would only take place if there was some selective pressure to change shape. However, the identification of morphospecies represented by multiple genotypes indicates where future work could examine potential morphological variability. As shown by Huber *et al.* (1997), morphological differences between *G. siphonifera* Types I and II were only observed with close scrutiny. In this study the only morphospecies found to have a large range of gross morphological variability was *T. quinqueloba*. Unfortunately, one section of the subarctic transect had a heterogeneous population, and sampling density within the second section of the subarctic transect was too low to determine whether the population was homogeneous (Fig. 4-11). There is a possibility that the *T. quinqueloba* Type IIb genotype has entirely sinistral coiling (Chapter 4, section 4.3.3), but this requires further investigation.

Foraminiferal genotype/morphotype relationships can be investigated by examining specimens by SEM prior to the amplification and sequencing of SSU rDNA. Holzmann and Pawlowski (1996) successfully used this approach with benthic foraminifers. They found that unless the foraminiferal specimens were examined under the SEM immediately prior to DNA extraction, the DNA amplification results were poor. However, the study had the benefit of being able to use freshly collected benthic foraminiferal specimens, as opposed to the time lag that would be associated with ship collected planktic foraminiferal specimens. As cruises are typically weeks long, the specimens would have to be frozen until SEM imaging was possible. The freezing and thawing of the specimens reduces their viability, and it would be hazardous to attempt this method prior to extraction.

As the morphological discrimination of genotypes within a morphospecies is likely to be difficult, are there any additional proxies that could be utilised ? The use of chemical proxies is the most probable option. Analysis of test chemistry could incorporate stable isotopes ($\delta^{18}O$ and $\delta^{13}C$), Mg/Ca ratios or Cd/Ca ratios. The increasing sensitivity of mass spectrometers would enable analysis of few, or even individual, specimens.

8.1.5. Speciation mechanisms

If the genotypic variation observed within planktic foraminiferal morphospecies is potentially indicative of cryptic speciation, what then are the possible mechanisms that could have caused the speciation events ? A number of speciation mechanisms have been described, though most are terrestrial models. The allopatric model suggests that speciation occurs due to geographic isolation (Mayr, 1942; Coyne, 1992). In this model natural selection or genetic drift would act on the isolated populations. However, planktic foraminiferal genotypes have a high dispersal potential as they mix globally (section 8.2) limiting potential isolation events in the present day. Alternatively, the sympatric model suggests that speciation can occur without effective geographic isolation (see Maynard Smith, 1966). Huber *et al.* (1997) suggested that divergence in the synchronicity of the reproductive cycle may be a mechanism by which sympatric speciation could take place (see below). Although it has been suggested that in the marine plankton the sympatric speciation model is likely (Lazarus, 1983; Lazarus *et al.*, 1995), the isolation of co-existing specimens seems problematic within the mixed layers of the water column.

Perhaps more suited to planktic foraminifers would be a model whereby an initial geographic isolation is followed by re-circulation, resulting in genotypes living together in sympatry. The initial isolation may be caused by ocean circulation changes induced by climatic cycling, or due to the discontinuous dispersal of specimens within the oceans from, for example, upwelling cells. Indeed, even in marine species with a high potential for dispersal, the division of populations has been observed (see the comprehensive review by Palumbi, 1994), which can result in genetically distinct populations. Whether re-mixing of the populations led to genetic recombination or reproductive isolation would depend on the degree and nature of the genetic divergence during isolation.

Reproduction in planktic foraminifers consists of the broadcast release of gametes (Hemleben *et al.*, 1989), therefore the synchroneity of gametogenesis is highly important (Palumbi, 1994; Huber *et al.*, 1997). In some spinose planktic foraminiferal morphospecies, gametogenesis has been shown to be related to distinct phases of the moon (Bijma *et al.*, 1990b). Any deviation from the "synchronised" time of reproduction could soon lead to divergent populations. Similarly, two closely related species of sea urchin were found to have reproductive cycles separated by a reproductive lag of 15 days (Lessios, 1984). Since the Isthmus of Panama separates these species, it is unclear whether they would be sexually compatible if they remixed. However, Lessios (1984) suggested that if they did remix, the reproductive lag would be sufficient to prevent them from reproducing, even if they were sexually compatible. At the molecular level, reproductive isolation could arise from the molecular isolation of the proteins controlling gamete recognition, which are known to have high evolution rates (Metz and Palumbi, 1996; Biermann, 1998). Evolution

of these proteins could result in gametes becoming incompatible resulting in reproductive isolation.

Speciation in planktic foraminifers is likely to be controlled by a combination of mechanisms, rather than any single mechanism. However, at present the limited knowledge of planktic foraminiferal biology restricts the understanding of the mechanisms that govern speciation and reproductive isolation within these organisms.

8.1.6. The implications of planktic foraminiferal cryptic speciation

The genetic diversity of planktic foraminifera in the North Atlantic is much higher than fossil record interpretations show. The evidence for cryptic speciation based on evolutionary distances within the 505 bp molecular phylogeny, the nucleotide sequence divergences within the variable regions of the SSU rRNA gene and the genotype distribution data is quite persuasive. Although some of the divergences are likely to be quite ancient, such as the divergence of the two extant *G. ruber* white lineage's (~ 22 Ma; Darling *et al.*, 1999), other divergences are likely to be considerably more recent. Assuming a constant evolution rate within the high-latitude morphospecies *G. bulloides* and *T. quinqueloba*, the relatively small divergences within the generalist Type II genotype clusters possibly reflect divergences within the Quaternary period.

As the diversity of planktic foraminifers is much higher than fossil record interpretations suggest, it is clear that the morphological species concept does not provide a complete picture. Instead, a combination of molecular, biological, morphological and chemical evidence will provide a more complete insight into the

evolutionary biology of planktic foraminifera, in particular the processes that govern speciation and extinction. In addition, it is possible that the morphological variation previously attributed to environment induced phenotypic (ecophenotypic) variation may actually reflect genotypic variation. Of course, if the genotypes were adapted to different environments, they would still reflect ecological variation within the oceans. It must also be noted that even though genotypes within a morphospecies may have become reproductively isolated, this does not necessarily imply that the genotypes have acquired different habitat adaptations since the genotypes may have been isolated in similar environments in geographically isolated areas. However, the genotypes within *G. bulloides*, *T. quinqueloba*, *G. siphonifera* and *N. pachyderma* do have different distributions and potentially different adaptations. The data shows the presence of generalist genotypes (*G. bulloides* Types IIa and IIb and *G. siphonifera* Types IIa and IIb) and specialist genotypes (*G. bulloides* Type Ib and *G. siphonifera* Type I). This is highly indicative of genotype habitat adaptation and has significant implications for palaeoceanographic and palaeoclimatic investigations, as it could potentially provide clarification and improved resolution in the fossil record interpretation. Environmental reconstructions that utilise planktic foraminiferal morphospecies must now take account of the level of genotype complexity that exists in the oceans. This is particularly important if the distinct genotypes have different habitat preferences, as shown by the distribution of *O. universa* genotypes and their relationship to ocean productivity (de Vargas *et al.*, 1999).

When planktic foraminiferal morphospecies are used in transfer function investigations it is assumed that each morphospecies assemblage in the sedimentary record represents a specific SST, and that the habitat of living planktic foraminiferal

specimens is representative of the equivalent foraminifers in the fossil record (see Chapter 1). However, morphospecies are now known to be commonly represented by more than one genotype, which may have different habitat preferences. This could result in bias, or even errors, being introduced into transfer function calculations.

Stable isotope based investigations also face potential inaccuracy. It has been shown that *G. siphonifera* Type I and Type II have different stable isotope signatures in the field, and in the laboratory they show different carbon and oxygen isotope fractionation (Bijma *et al.*, 1998). They suggested that the difference in the $\delta^{13}C$ between Type I and Type II could be used as a productivity indicator, since the $\delta^{13}C$ of Type I gets lighter at higher feeding frequency but Type II remains unaffected (Bijma *et al.*, 1998). However, since *G. siphonifera* Types I and II were difficult to distinguish in the fossil record, this limited their ability to interpret the *G. siphonifera* isotope record (Bijma *et al.*, 1998). In other words, if *G. siphonifera* Types I and II are not differentiated prior to isotopic analysis, two different stable isotope signatures are combined thus reducing the resolution of the isotopic record.

Similarly, this is also important when morphospecies are utilised for geochemical investigations, such as in the case of *G. bulloides* and *G. ruber*, which are each represented by multiple genotypes. Although entirely speculative at this point, there is the potential that the genotypes of *G. bulloides* and *G. ruber* morphospecies have different stable isotope signatures due to either occupying different habitats within the water column, or due to having divergent biological adaptations. If their genotypes do have different isotopic signatures, then differentiating them provides a potential source of additional geochemical information.

The presence of cryptic species in the North Atlantic should not be taken as being negative for palaeoceanographic investigations, but rather as an opportunity to improve the resolution of the fossil record. Further investigation will be required to (a) examine seasonal variations in genotype distribution, (b) determine the complete genotype populations, (c) determine any genotype/morphotype link, (d) investigate further chemical proxies. Perhaps by combining culture experiments with genotyping, the relationship between morphology, genotype and environment may be better understood. If the genotypes could be identified within the marine sediments they would provide new, higher resolution proxies for palaeoceanographic investigation.

8.2. Gene flow

Many planktic foraminiferal morphospecies are cosmopolitan, being found in similar habitats in different oceans (Bé, 1977). The distribution of planktic foraminiferal morphospecies within the modern oceans is represented in simple terms by five faunal provinces, each with a characteristic species assemblage, which are mirror imaged between hemispheres (Fig. 1-1). The question of whether genetic isolation or genetic interchange occurs within planktic foraminiferal morphospecies from different oceans of the globe, can be investigated by comparing within morphospecies genotypic variation. Previous investigation has compared genotype populations from the Coral Sea, Caribbean, and the Southern Californian Bight (Darling *et al.*, 1999). Comparison of *O. universa* genotype populations from the Atlantic Ocean has shown that their distribution is correlated to specific hydrographic provinces (de Vargas *et al.*, 1999). As the presence of specific genotypes are recognised in different regions of the oceans it becomes possible to determine the extent of genetic interchange occurring between planktic foraminiferal populations utilising planktic foraminiferal morphospecies specifically from the North Atlantic. There are two sections to this discussion. The first is a discussion of bipolarity within the Atlantic Ocean, and the second is a discussion of genetic exchange between the Atlantic and Indo-Pacific Oceans.

8.2.1. Bipolarity/antitropicality

Bipolar species distribution, or antitropical distribution as it is often known as, refers to a discontinuous distribution of cold water taxa within the oceans, separated by the tropical, low latitude, water masses. Bipolarity has posed one of the

great enigmas within foraminiferal research (Lipps, 1979), and has also been observed in many marine invertebrate, vertebrate, and plant groups. A number of mechanisms have been proposed for the antitropical distribution of marine taxa (see review by Lindberg, 1991). The mechanisms for creating discontinuous distributions can be broken down into two main categories: vicariance and dispersal. The vicariance theory proposes that an original cool water population was separated into a northern and southern component by the formation of the tropical province. The dispersal mechanism proposed that genetic exchange occurred from one polar region to the other. Whether the gene flow continues remains to be determined.

The use of molecular phylogenetic data provides a way in which the question of genetic isolation versus genetic exchange can be addressed. Three predominantly bipolar planktic foraminiferal morphospecies have been investigated in this study (Chapters 4 and 7). The subpolar representatives from the North Atlantic Ocean (this study) can now be compared with their equivalents from the subantarctic region of the South Atlantic (Darling *et al.*, submitted). If the subarctic and subantarctic populations have been isolated from one another, the genotypes from each region would be expected to be genetically distinct. Conversely, if regular or continual genetic exchange has been occurring between these regions, the genotypes would be expected to have nucleotide sequence homogeneity.

8.2.1.1. Phylogenetic relationship between bipolar planktic foraminiferal morphospecies

The bipolar morphospecies examined were *Neogloboquadrina pachyderma* (dextral), *Globigerina bulloides* and *Turborotalita quinqueloba* collected from the

subarctic Atlantic (this study), and from the subantarctic Atlantic (between the Falkland Islands and the Antarctic peninsula) (Darling *et al.*, submitted). The foraminiferal molecular phylogeny, with the bipolar genotypes highlighted, is shown in Fig. 8-1.

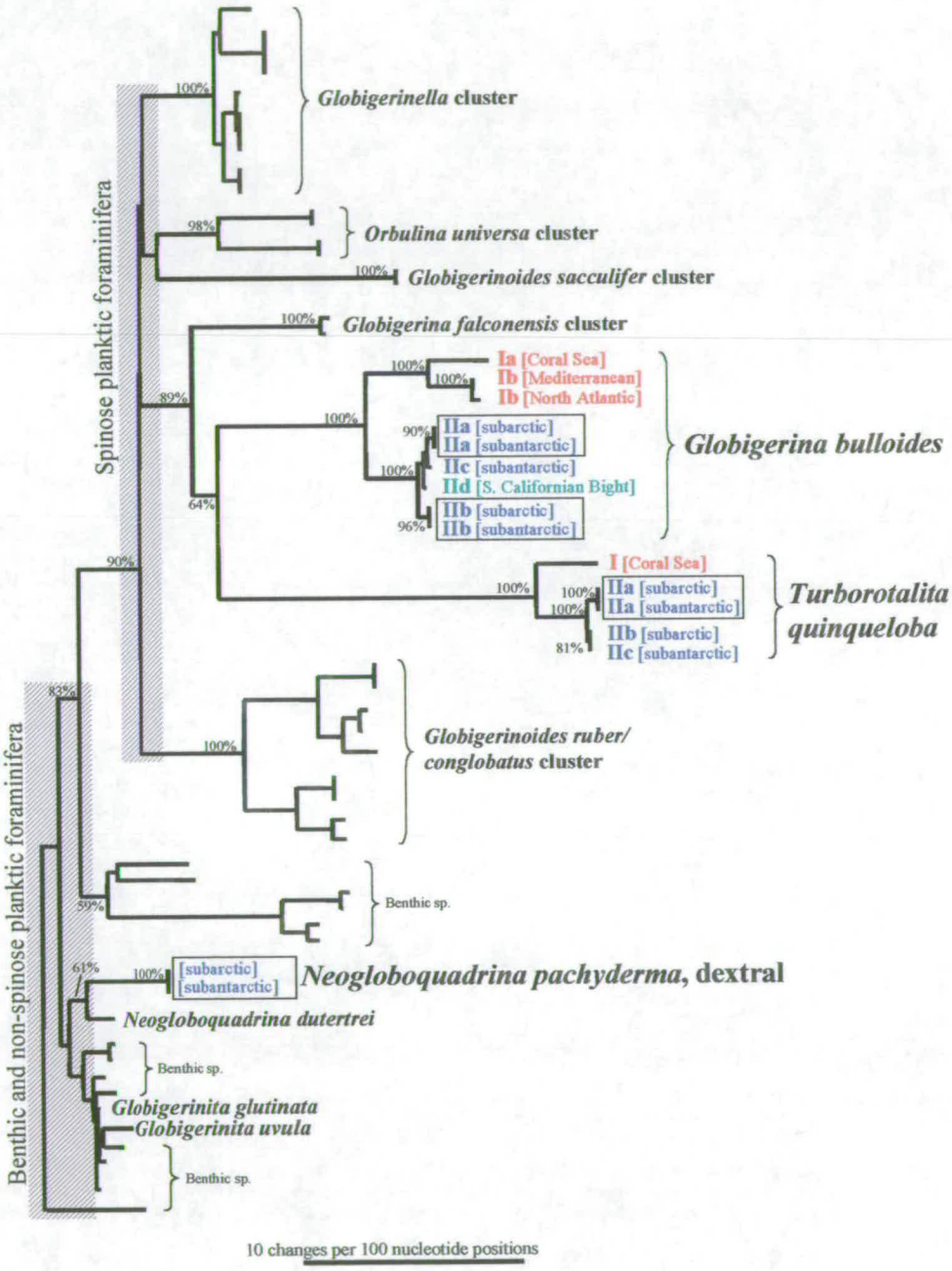


Fig. 8-1. Neighbour-joining phylogeny based on 505 unambiguously aligned sites, highlighting the bipolar morphospecies. The genotypes that are identical in both subpolar regions are boxed. The genotypes in blue, green and red represent those found in cool, transitional and warm waters respectively.

The molecular phylogeny (Fig. 8-1) shows that all three bipolar morphospecies examined have at least one genotype in each subpolar region which is identical in all nucleotide sites (505 bp) used for analysis. Manual alignment (see appendices) also shows that they are identical throughout the entire SSU rDNA fragment. The identical genotypes are *G. bulloides* Type IIa and Type IIb, *T. quinqueloba* Type IIa, and *N. pachyderma* (D).

The data is indicative of genetic interchange between the Arctic and Antarctic subpolar populations. Four identical genotypes have been identified from the Arctic and Antarctic subpolar regions of the Atlantic, showing that genetic exchange has occurred relatively recently, or even continuously.

8.2.1.2. Mechanism of genetic exchange between subpolar populations

It is evident that genetic exchange between the Arctic and Antarctic subpolar populations has occurred. The Atlantic Ocean circulation and faunal provinces (after Bé, 1977) shown in Fig. 8-2 demonstrate the distances involved, provinces to be crossed and surface currents to be negotiated for genetic exchange to take place between the subpolar regions.

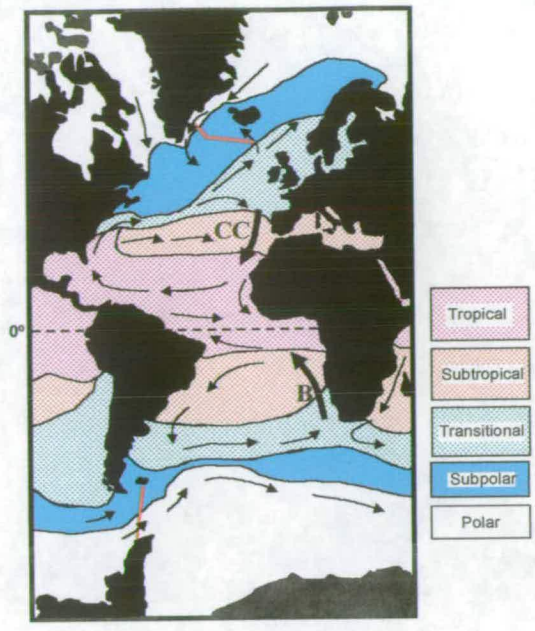


Fig. 8-2. Atlantic Ocean surface circulation, and faunal provinces (after Bé, 1977). The cool eastern boundary currents are highlighted, with **CC** denoting the Canary Current, and **B** the Benguela Current. The location of the collection transects are denoted as red lines.

The three morphospecies examined display a predominantly antitropical distribution (Bé and Tolderlund, 1971; Bé, 1977), although they are also known to reside in upwelling cells within the lower latitudes (Ufkes and Zachariasse, 1993; Boltovskoy *et al.*, 1996; Ufkes *et al.*, 1998), where upwelling produces lower SST's than the surrounding area. For genetic exchange to occur, the cooler water genotypes would need to traverse the warm tropical water mass by negotiating the equatorial current systems. Passage does not appear to be occurring across the open-ocean equatorial region at present since sediment core tops from the central equatorial Atlantic indicate that *G. bulloides*, *N. pachyderma* (D) and *T. quinqueloba* are not present in this region (McIntyre *et al.*, 1989). Rather, *G. bulloides* and *N. pachyderma* (no data for *T. quinqueloba*) occur in "pulses" throughout the recent past (Ericson and Wollin, 1956). Although the timing of the *G. bulloides* and *N.*

pachyderma pulses has not been clarified, they suggest that climatic cycling is the likely factor influencing when they transit the tropical waters. Indeed, during the Last Glacial Maximum, a more subpolar assemblage including *G. bulloides* and *N. pachyderma* (D) dominated the equatorial eastern South Atlantic (McIntyre *et al.*, 1989).

Plankton tows have shown that along the west African margin at present, "cool water" morphospecies, including *G. bulloides*, *T. quinqueloba* and *N. pachyderma* (D), penetrate into the low latitude regions of the North Atlantic (Thiede, 1975; Cifelli and Bernier, 1976). Further, during the glacial periods of the Quaternary, the Atlantic eastern boundary surface currents (Canary Current and Benguela Current, see Fig. 8-2) are thought to have strengthened (Mix and Morey, 1996; Zhao *et al.*, 1995), enhancing the transport of cool water morphospecies into tropical areas. As mentioned above, upwelling cells within the low latitudes contain planktic foraminiferal morphospecies that are typical of high latitude, cool water assemblages. Examination of the marine sediments has shown that upwelling intensified during glacial periods in both the subtropical North Atlantic (Meggers *et al.*, 1998) and the subtropical South Atlantic (Lohmann, 1992; Little *et al.*, 1997) eastern boundary current systems. This, combined with the strengthening of the eastern boundary currents, promoted the "invasion" of high latitude planktic foraminiferal species into the tropics, as observed in the sedimentary record (McIntyre *et al.*, 1989).

Whether enhanced tropical penetration of cool water genotypes would be sufficient for the complete transit across the tropical waters is unclear, since it is a formidable barrier and planktic foraminifers do not encyst which would otherwise

provide protection to inhospitable environments. Recent research however, has proposed an ice age cooling of the tropics by up to 6°C from present day sea surface temperatures (SST) (Mix *et al.*, 1999). This could mean that during glacial periods the eastern tropical Atlantic SST could have been as low as ~20°C (adjusted from Levitus and Boyer, 1994), which is far more hospitable to "cool water" genotypes than at present. In addition to the lower tropical SST during glacial periods, a further mechanism by which the cool water foraminiferal species might transit inhospitable waters is by tropical submergence, whereby the planktic foraminifers would sink to a depth that is within their thermal tolerance limits and traverse the tropics *via* the equatorial undercurrents. Similarly, Boltovskoy *et al.* (1996) suggested that the subpolar/polar species observed in low latitude upwelling cells, along the eastern margin of South America, are also indicative of submergence. They proposed that the "cold water" species present within the upwelling cells were connected to the subpolar/polar region not across the surface waters but *via* a route at 400-500 m depth where the SST was suitably low.

It is likely that the low latitude upwelling cells maintain cool water foraminiferal populations until the opportunity to transit the inhospitable tropics becomes possible during glacial periods. It must be noted that solely traversing the tropics may not be sufficient for genetic mixing to take place. For genetic exchange to occur between the Arctic and Antarctic subpolar regions, the foraminifera would not only have to traverse the tropics but also complete a circuit of the North and South Atlantic surface current gyral system. Norris (1999) argued that the dispersal capability of planktic foraminiferal species is extraordinary. Indeed, the investigation of genotype distribution in the North Atlantic has shown that *G. bulloides* Type IIb is

also present in the subtropical Azores Current (Fig. 4-7). The SST was 19.3°C, which is very similar to the estimated ice age tropical SST (Mix *et al.*, 1999), showing that such a tropical SST would be sufficiently low for *G. bulloides* Type IIb to tolerate. Its presence in the Azores Current indicates that this genotype may have completed a circuit of the North Atlantic gyre. The occurrence of this genotype in the subtropics could also be explained by it inhabiting the waters associated with upwelling at the Azores frontal zone. Alternatively, this individual has arrived from an Atlantic margin upwelling cell.

A key to understanding the role the Atlantic upwelling regions have in genetic exchange between the Arctic and Antarctic subpolar populations is to determine which, if any, of the North Atlantic genotypes identified in this study represent the planktic foraminiferal morphospecies in the upwelling zones. Further investigation of genotype distribution within the Atlantic Ocean is required to fully understand the extent to which genetic exchange may be occurring between Atlantic bipolar genotypes at present. In addition, if it were possible to determine a link between genotype and morphotype, then genetic exchange between the bipolar populations could be examined in more detail within the deep-sea sediment record. The exact timing of the genetic interchange events could then be investigated.

8.2.2. Genetic exchange between North Atlantic and Pacific transitional/ subtropical planktic foraminifer populations

The majority of planktic foraminiferal morphospecies have representatives in similar environments in each of the major oceans. It has been shown that genetic exchange occurs between bipolar populations within the Atlantic (see section 8.2.1.1). A similar approach can be used to determine whether transitional/ subtropical planktic foraminiferal populations from the North Atlantic and Pacific Oceans mix.

To examine genetic exchange between the North Atlantic and the Pacific, genotypes of *G. ruber, G. siphonifera, G. bulloides* and *G. falconensis* from the North Atlantic (this study), Southern Californian Bight and Coral Sea (Darling *et al.*, 1996, 1997, 1999) have been compared. By combining the information from different morphospecies, further insight into the way that genetic exchange occurs between populations from different oceans can be obtained.

The molecular phylogenetic relationships within these planktic foraminiferal morphospecies have been discussed in Chapters 4, 5, and 6. A number of genotypes from the North Atlantic and the Pacific have very close if not identical evolutionary relationships. Yet, in contrast, there are also a number of genotypes from the North Atlantic and the Pacific which are quite divergent from one another.

8.2.2.1. The degree of evolutionary relatedness between North Atlantic and Pacific Ocean genotypes

(A) A close evolutionary relationship has been observed within *G. siphonifera* Types IIa and IIb and *G. ruber* white Type I and Type II (see Fig. 8-3).

Comparison of the genotypes obtained from the North Atlantic, Southern Californian Bight and Coral Sea shows that:

(a) The *G. siphonifera* Type IIa(3) and Type IIa(4) genotypes, from the North Atlantic and Southern Californian Bight respectively, are separated by only 2 base differences within the variable regions of the SSU rDNA fragment (Fig. 8-3A).

(b) The *G. siphonifera* Type IIb genotypes from the North Atlantic and Southern Californian Bight are identical throughout the entire ~1000 bp SSU rDNA fragment (Fig. 8-3B).

(c) The *G. ruber* white Type II genotypes from the North Atlantic and Southern Californian Bight are separated by only 1 base substitution throughout the entire ~1000 bp SSU rDNA fragment (Fig. 8-3C).

(d) The *G. ruber* white Type I genotypes obtained from the North Atlantic and Coral Sea are separated by only 2 base substitutions throughout the entire ~1000 bp SSU rDNA fragment (Fig. 8-3D).

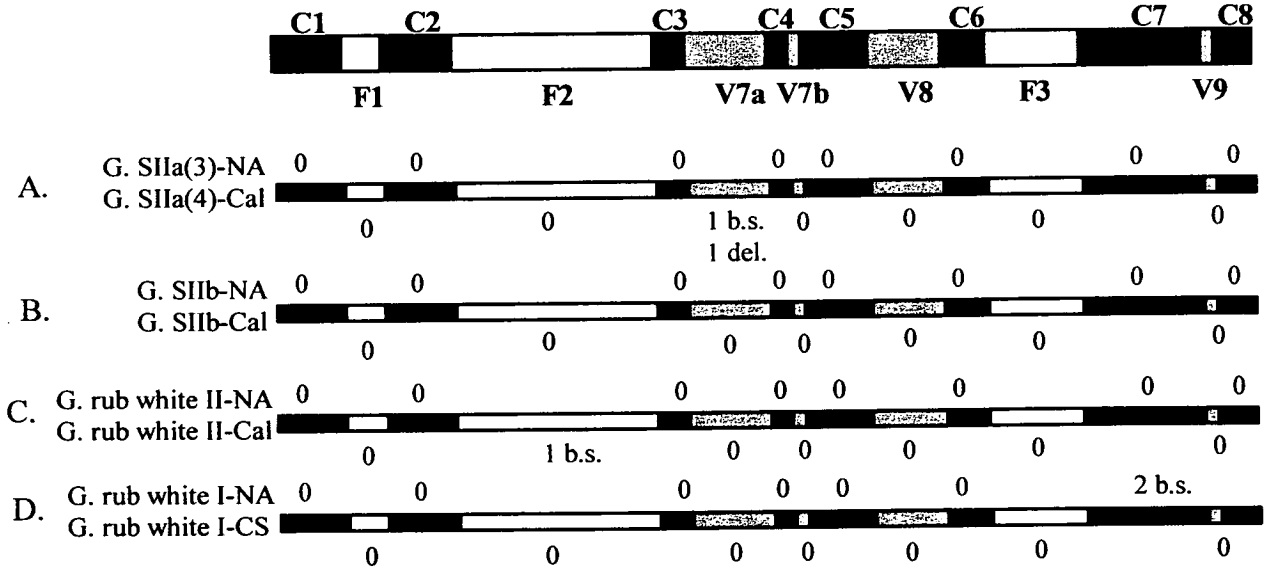


Fig. 8-3. Comparison of genotypes using partial SSU rRNA gene fragments. NA denotes the North Atlantic, Cal the Southern Californian Bight, CS the Coral Sea, bs denotes a base substitution difference and del denotes a base deletion/insertion difference.

These data suggests that there has been recent, or possibly even constant, genetic exchange between these North Atlantic and Pacific Ocean genotype populations.

(B) In contrast, divergences have been observed between genotypes of *G. falconensis*, *G. bulloides* Type I and Type II, and *G. siphonifera* Type IIa. The level of divergence is far higher in *G. falconensis* and *G. bulloides* Type I than in *G. siphonifera* Type IIa. The degree of genetic divergence does not necessarily mean that the genotypes are isolated spatially. It is quite possible that the Pacific genotypes will be found in the Atlantic and *vice versa*.

(a) The North Atlantic and Coral Sea *G. falconensis* genotypes are quite divergent when the entire ~1000 bp SSU rDNA fragments are compared. In addition, when the low evolution rate observed within this lineage is taken into consideration, such changes suggest that the two populations may not have mixed genetically for a long period of time (Fig. 8-4A).

(b) The North Atlantic and Coral Sea *G. bulloides* Type I genotypes are also quite divergent (4.7 % within the 505 bp phylogeny; see Chapter 4, section 4.2.1). Although the higher evolutionary distance reflects in part a higher evolution rate within this lineage compared with *G. falconensis*, it also indicates that a substantial period of time may have passed since the populations mixed genetically.

(c) The *G. siphonifera* Type IIa(3) genotype from the North Atlantic is quite divergent from the Coral Sea Type IIa(2) genotype when the entire ~1000 bp SSU rDNA fragments are compared (Fig. 8-4B). This shows that a longer period of time has elapsed since the populations genetically mixed, than is evident between the

North Atlantic Type IIa(3) and the S. Californian Bight Type IIa(4) genotypes (Fig. 8-3A).

(d) The *G. bulloides* Type IId genotype from the Southern Californian Bight is quite divergent from Type IIb in the North Atlantic (Fig. 8-4C).

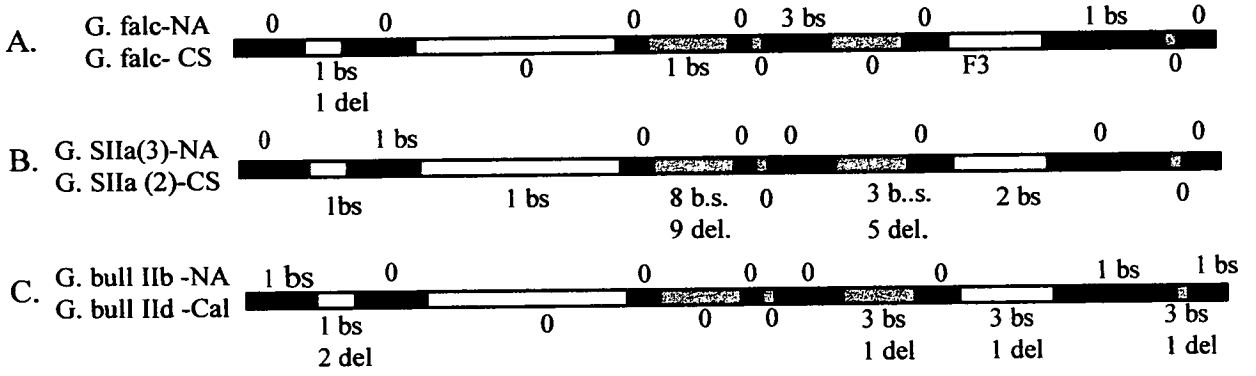


Fig. 8-4. Comparison of genotypes using partial SSU rRNA gene fragments. Abbreviations are the same as in Fig. 8-3. Within the *G. falconensis* comparison, F3 denotes a region with 99 aligned sites which have no base differences, and a region that is unalignable which is 60 bases and 48 bases in length for the North Atlantic and Coral Sea genotypes respectively.

8.2.2.2. What are the possible routes into and out of the North Atlantic ?

It is apparent that genetic interchange does occur between the Atlantic and Pacific Oceans. The trans-Arctic route (Vermeij, 1991) is far too inhospitable for transitional/subtropical planktic foraminiferal species. There are two possible alternative routes of transit for the Atlantic and Indo-Pacific populations. The first is *via* the Drake Passage, and the second is *via* the Cape of Good Hope, South Africa (Fig. 8-5). The major surface currents of the present day oceans are shown in Fig. 8-5. Although a simplification, it serves to illustrate the direction of surface water flow within the large gyres that occupy the Atlantic, Pacific, and Indian Oceans, in addition to the other important surface currents of the oceans.

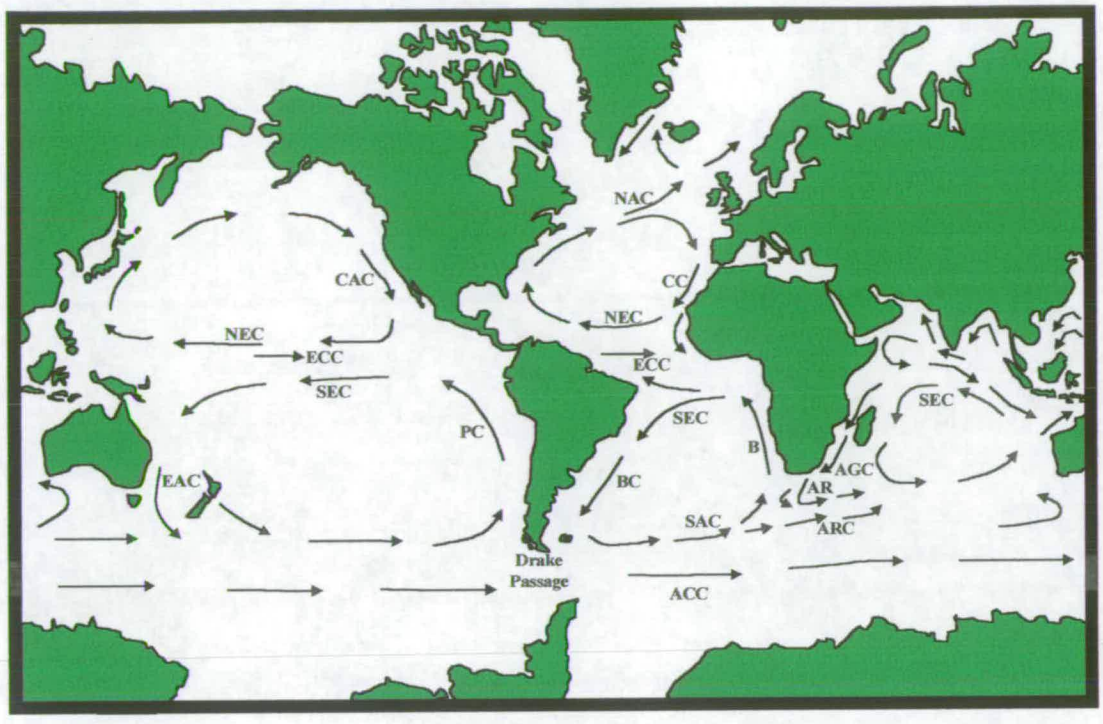


Fig. 8-5. Global ocean surface circulation. ACC denotes the Antarctic Circumpolar Current, AGC the Aghulas Current, AR the Aghulas Retroflection, ARC the Agulhas Return Current, B the Benguela Current, BC the Brazil Current, CAC the Californian Current, CC the Canary Current, EAC the East Australian Current, ECC Equatorial Countercurrents, NAC the North Atlantic Current, NEC the North Equatorial Current, PC the Peru Current, SAC the South Atlantic Current, SEC the South Equatorial Current. Agulhas region oceanography adapted from Boebel *et al.* (1998).

If gene flow was occurring between the Pacific and Atlantic Oceans *via* the Drake Passage, it would occur from west to east due to the prevailing surface currents (Fig. 8-5). However, the Antarctic Circumpolar Current (ACC) provides an extremely inhospitable habitat for transitional/subtropical morphospecies such as *G. siphonifera, G. ruber* and *G. falconensis*, since the sea surface temperature (SST) ranges from approximately 3.5°C (winter) to 5.5°C (summer) (Boltovskoy, 1971). In addition, plankton tows from the Drake Passage during the summer, when SSTs are at their highest and therefore when gene flow is most likely to occur, shows that no transitional/subtropical planktic foraminiferal morphospecies exist within the water

column. Indeed, in the South Atlantic the transitional/subtropical species first occur much further north than the ACC waters derived from the Pacific (Boltovskoy *et al.*, 1996). This shows that transitional/subtropical planktic foraminiferal morphospecies do not transit from the Pacific to the Atlantic *via* the Drake Passage.

Previous investigations (Darling *et al.*, 1997, 1999) have shown that Caribbean and Coral Sea populations of *Globigerinoides sacculifer* and *Orbulina universa* are genetically homogeneous. They suggested that the genetic exchange within the tropical/subtropical regions of the Atlantic, Pacific and Indian Oceans is relatively unrestricted and that, since the surface currents in these regions flow in a predominantly westerly direction, genotype transit most likely occurs from the Indo-Pacific to the Atlantic Ocean *via* the Cape of South Africa (Darling *et al.*, 1999). In addition, it was apparent that populations of *O. universa*, *G. ruber* and *G. bulloides* from the Southern Californian Bight and the Coral Sea were genetically isolated from one another (Darling *et al.*, 1999).

As genetic exchange clearly occurs between the Atlantic and Pacific Oceans, probably *via* the Cape of South Africa, how could this occur ? The physical oceanography of this region is characterised by the westward flowing Agulhas Current, and the eastward flowing South Atlantic Current, Agulhas Return Current, and South Indian Ocean Current (Fig. 8-5) which enables water to be exchanged between the Atlantic and Indian Oceans (Lutjeharms, 1996). The water temperature in this ocean region ranges from >15°C in winter to >20°C in summer (Levitus and Boyer, 1994). This is relatively warm and many planktic foraminifer morphospecies could tolerate this, especially the more transitional morphospecies. However, this water may be insufficiently warm for tropical specialists to tolerate during transit. In

addition, the Agulhas Current causes warm ring cores (cells of warm Indian Ocean water), to spin off into the Atlantic. These warm rings can be hundreds of kilometres in diameter, persist for 3 to 4 years, and transit as far north as 25°S (see review by Lutjeharms, 1996). These rings provide a mechanism within which planktic organisms, including foraminifers, may be transported from the Indian Ocean to the low latitudes of the Atlantic Ocean.

Although the ocean surface currents allow transit of planktic foraminifers in both directions, the dominant transport of planktic foraminifers is likely to be westward from the Indian to the Atlantic Ocean. This is due to the dominance of the westward flowing Agulhas Current and the warm ring cores which would provide a habitat within which the tropical/subtropical planktic foraminiferal species could survive.

8.2.2.3. Species tolerance to movement between the Pacific and Atlantic Oceans

Although most planktic foraminiferal species occur in similar environments in each of the oceans, there are a number of subtropical/tropical species that have become extinct from either the Atlantic Ocean or the Indo-Pacific. In the Atlantic *Globorotaloides hexagona*, *Globoquadrina conglomerata* and *Globigerinella adamsi* are extinct (Bé and Tolderlund, 1971; Bé, 1977) and in the Indo-Pacific *Globigerinoides ruber* (pink) is extinct (Thompson *et al.*, 1979). Why then are these species extinct in the Atlantic or the Pacific Oceans when it is clear that planktic foraminifers have a route of transit between the Atlantic and Indo-Pacific ? Each of the species extinct in one ocean has a clear preference for warm waters suggesting that they may not have repopulated the Pacific or Atlantic due to their inability to

survive transit between the Atlantic and Indo-Pacific. Norris (1999) also has suggested that if they do tolerate transit they may not be able find a suitable habitat to reproduce once they reach the other ocean. For example; it is known that within the subtropical/tropical North Atlantic *G. ruber* (pink) prefers warm (summer) waters (Deuser and Ross, 1989; Pujol and Grazzini, 1995; Ganssen and Kroon, in press), hence may not tolerate the cool South Atlantic Current waters encountered in transit between the Atlantic and Indian Ocean.

Other factors that control the ability of foraminiferal genotypes to transit the oceans include reproductive turnover and potential lifetime. Some planktic foraminifera reproduce monthly, although others may be dependent upon feeding and maturity (Hemleben *et al.*, 1989; Bijma *et al.*, 1990b). Assuming a constant surface flow rate of 0.56 ms^{-1} (this is the average speed of a warm ring core in the South Atlantic; Boebel *et al.*, 1998), the distance a single planktic foraminiferal specimen could transit during one month is approximately 1350 km. Of course, this distance will be extremely variable depending upon the velocity of the surface currents. Their transit range would be considerably further if specimens were able to reproduce *en route*, extending their effective dispersal range. Obviously, reproduction could only occur if the ocean conditions were favourable. Culture experiments by Bijma *et al.* (1990a) showed that the temperature tolerances of planktic foraminiferal species closely corresponded to their distribution in the modern ocean. However, they showed that the survival time and the percentage of specimens that undergo gametogenesis varies considerably with temperature (salinity ranges found in the ocean do not appear to be a limiting factor for the species studied). Therefore, unfavourable SST conditions reduce the survival time and the percentage of

specimens that undergo gametogenesis, thus reducing the potential transit range of planktic foraminifers. In addition, if a foraminiferal specimen is adapted to live in eutrophic waters then it is unlikely to survive in oligotrophic areas, also reducing its potential transit range. It is known that some planktic foraminifer species can survive in extremely hostile conditions, such as *N. pachyderma* (S) living in polar sea ice without any detrimental effect (Hemleben *et al.*, 1989). Further, it is possible that some planktic foraminifers are adapted to survive between seasons by staying in a dormant state in the water column until the next season, or until there is a phytoplankton bloom. If specimens were in a dormant state, then this could aid transit of the oceans.

8.2.2.4. Habitat tolerance

It is evident from the molecular data presented above, that genetic exchange between the North Atlantic and the Pacific appears to be commonplace for a number of planktic foraminiferal genotypes. Gene flow has been shown to occur between populations of *G. siphonifera* (Types IIa and IIb) and *G. ruber* white (Type II) from the North Atlantic and the Southern Californian Bight. Similarly, gene flow has occurred between populations from the North Atlantic and the Coral Sea since genotypes of *G. sacculifer* and *O. universa* from the Coral Sea and the Caribbean are genetically homogeneous (Darling *et al.*, 1999), and *G. ruber* white Type I from the North Atlantic (this study) and Coral Sea (Darling *et al*, 1999) have a very close evolutionary relationship. Conversely however, there are some North Atlantic – Coral Sea/Southern Californian Bight genotypes which are more genetically divergent. Whether or not the *G. falconensis* and *G. bulloides* Type I genotypes

transit the Cape of South Africa is unknown. Both of these morphospecies are found in the Southeast Atlantic and also in the Indian Ocean. *Globigerina falconensis* is common in the south Atlantic subtropical gyre where it accounts for 20 % of the total assemblage (Niebler and Gersonde, 1998), and it can also account for 8.4 % of the total assemblage in the Benguela Current system (Fig. 8-5) (Giraudeau, 1993). The SST in the Cape of South Africa region ranges from ~ 15-20°C (Levitus and Boyer, 1994), although there are stretches of near-shore water where the mean annual SST is 13°C due to coastal upwelling (Giraudeau, 1993). The *G. falconensis* genotype was found in transitional and subtropical waters of the North Atlantic in water temperatures as low as ~13°C (Chapter 4), suggesting that this genotype would be able to tolerate transit round the Cape of South Africa. However, it is not known as yet whether *G. falconensis* in the South Atlantic or the South Indian Ocean is represented by the same genotype as in the North Atlantic. An investigation of genotypes within these regions would aid the understanding of genetic exchange between Atlantic and Indo-Pacific *G. falconensis* populations. The distribution of *G. bulloides* Type I is also unknown within these regions. Although this morphospecies is very abundant (up to 45 % of the assemblage) in the Benguela Current system of the South Atlantic (Giraudeau, 1993), it is not known which genotype represents *G. bulloides* in this region. Interpretation is again limited until the Benguela Current and Indian Ocean genotypes have been investigated. It is unknown at present whether all genotypes occur in both oceans. It is likely that genotype habitat tolerance controls the extent of genotype distribution within the oceans.

Habitat tolerance is most likely to determine whether planktic foraminifers become isolated in regions of the oceans during climatic cycling. There are several

extant planktic foraminiferal morphospecies that are not present in the Atlantic or Pacific yet present day ocean circulation would permit transit into these oceans (Fig. 8-5). Therefore it is likely that either they cannot tolerate the habitats encountered during transit, or if they do tolerate transit they do not find a suitable habitat to reproduce (Norris, 1999). Whatever the cause, these morphospecies have not repopulated the ocean in which they became extinct. It is possible that specific genotypes will display a similar pattern, where some genotypes are tolerant of transit between the oceans and are able to reproduce once they get there, and others are not. The genotype distribution data from the North Atlantic shows that, in some cases, genotypes within a morphospecies have differing distribution patterns from one another. Additional investigation will be required to clarify their temporal and spatial differences. However, if genotypes do have habitat preferences, evidence suggests that this will most likely be linked to their habitat tolerance. Culture experiments by Bijma *et al.* (1990a) showed that the temperature tolerances of planktic foraminifera under laboratory conditions were consistent with their geographic distribution in the modern ocean. It was also shown by Bijma *et al.* (1990a) that the salinity tolerance of the species investigated exceeded the range found in the modern oceans. Salinity would not apparently limit the distribution of the spinose species investigated. As the geographic distribution of planktic foraminifers in the modern ocean is a good indicator of their temperature tolerance, it is clear that the distribution of genotypes would provide an indicator of genotype temperature tolerances.

8.2.2.5. Potential scenarios for global isolation and speciation

The *G. siphonifera* Type IIa(2) genotype from the Coral Sea shows considerable sequence divergence, within the variable regions of the SSU rRNA gene, from the N. Atlantic Type IIa(3) genotype (Fig. 8-4) and from the S. Californian Bight Type IIa(4) genotype. The glacial/interglacial cycling of the Quaternary may have resulted in the isolation of these *G. siphonifera* Type II genotypes, allowing them to become sufficiently distinct from one another to maintain their reproductive isolation. Such re-mixing of reproductively isolated genotypes may also explain why there is a degree of sequence divergence between the *G. siphonifera* Caribbean Type IIa(1) and N. Atlantic Type IIa(3) genotypes (Fig. 5-4). It is likely that a similar scenario resulted in the variable region sequence divergence observed between *G. bulloides* Type IId (Southern Californian Bight) and Type IIb (Atlantic).

The greater genetic divergences between the North Atlantic and Coral Sea *G. falconensis* genotypes and between the North Atlantic and Coral Sea *G. bulloides* Type I genotypes are indicative of longer periods of time following isolation. The relatively slow evolution rate within the *G. falconensis* lineage combined with the sequence divergence between the genotypes, especially within the foraminiferal specific insertion F3 (Fig. 8-4), suggests that the genotypes may have diverged prior to the Quaternary period. The *G. bulloides* Type I genotypes have a much larger genetic distance than the *G. falconensis* genotypes within a 505 bp molecular phylogeny (Chapter 4) which may be in part due to the higher evolution rate within the *G. bulloides* lineage (Chapter 3, section 3.4.1). It also suggests that the divergences may have occurred prior to the Quaternary period.

If the divergences arose prior to the Quaternary, how could this occur ? It possible that the genetic divergences could be a result of the cessation of direct ocean circulation between the tropical regions of the Atlantic and Pacific Oceans. There are two possible scenarios. First, it is possible that these genotypes stopped mixing with the tectonic uplift at the Isthmus of Panama, which closed the Central American Seaway and therefore the link between the tropical regions of the Atlantic and Pacific Oceans. Complete closure of the seaway is thought to have occurred between ~3.5-3.1 Ma (Saito, 1976; Keigwin, 1978, 1982), although it has been proposed that marine flooding has occurred several times across the Isthmus between 3-2 Ma (Cronin and Dowsett, 1996) which could have permitted genetic exchange between the genotypes. The closure would have prevented the genotypes mixing for the last 2 Ma, and could account for the level of sequence divergence between these genotypes. This scenario is supported by the substantial reproductive isolation observed in other taxa separated by the Isthmus of Panama, e.g. two genera of sea urchin (Lessios, 1984; Lessios and Cunningham, 1990) and a genus of snapping shrimp (Knowlton *et al.*, 1993). The second scenario is considerably more ancient than the 3.5 or 2-3 Ma datum. Indeed, investigation of the snapping shrimp genus *Alpheus*, by Knowlton *et al.* (1993), revealed that the divergences between some of the other Pacific and Caribbean taxa occurred several million years earlier. Knowlton *et al.* (1993) suggested that this was due to a change in ocean circulation patterns which initially stopped flow across the Central American Seaway between 12.9 and 7 Ma, followed by a very shallow water connection between the oceans (Duque-Caro, 1990). The shallow water connection may not have been sufficiently deep for planktic foraminifers to transit. Such a scenario could account for the larger genetic

divergences between the *G. bulloides* Type I cluster from the Type II cluster and between the *G. siphonifera* Type I genotype from the Type II cluster. Indeed, the molecular phylogeny indicates that within the *Globigerinella* lineage *G. siphonifera* Type I, *G. siphonifera* Type II and *G. calida* diverged over a short period of time. The molecular phylogeny also indicated that *G. calida* may have been in existence for 10-11 Ma. This datum is similar to when the ocean circulation between the Pacific and Atlantic initially ceased across the Central American Seaway (12.9-7 Ma), and may perhaps have been the driving force behind their speciation.

8.2.2.6. Conclusions

As yet, it is far too early to draw conclusions from the molecular phylogenetic data with regard to gene flow in the oceans. Additional genotypes must remain to be discovered within the oceans, which will alter the apparent trends shown by the limited data at present. In addition, the variations in evolution rate between, and within, morphospecies limits the use of the molecular phylogeny as a "clock" to provide a means with which to examine the timing of divergences. Gene flow between genotype populations within the oceans appears to be common. It has been shown that gene flow has occurred recently, or maybe even continues today, between subarctic and subantarctic genotypes within the Atlantic and also between genotypes from the North Atlantic and the Pacific Oceans. Habitat tolerance and preference will influence gene flow by affecting a genotype's ability to transit between the oceans, or by preventing it reproducing in unfavourable habitats.

8.3. Implications and further work

The molecular phylogenetic data and genotype distributions from the North Atlantic have provided some interesting information for planktic foraminiferal research. However, there are a number of key areas where the implications must be considered further. This section will address these issues and suggest possible ways in which they may be resolved or applied within palaeoceanographic investigations.

8.3.1. Taxonomy and phylogenetic implications

Arguably the most significant finding during this research has been the extent to which genotypic variation exists within morphologically defined planktic foraminiferal species in the North Atlantic. As previously discussed, the genotypic variation within morphospecies (combined with the genotype distributions) is highly indicative of cryptic speciation. This illustrates that the species concept based on morphology alone does not provide the complete picture. Indeed, biological diversity in planktic foraminifers must be significantly higher than previous morphological investigations have indicated. Further, it is possible that morphological variation attributed to phenotypic variability (i.e. the range of plasticity within a population or environmentally induced variation) may actually reflect genotypic variation.

Even though a taxonomic revision of some North Atlantic planktic foraminiferal morphospecies is required, there are a number of difficulties with applying the molecular data to mainstream foraminiferal research. First, what constitutes a distinct species ? As there is no specific evolutionary distance that defines a species level distinction we must look to other supportive evidence such as genotype distribution, biology or physiology.

Second, can we use the SSU rDNA molecular phylogeny to determine when evolutionary divergences took place ? There is considerable variation in evolution rates of the SSU rRNA gene between and within planktic foraminiferal morphospecies. This variability limits the use of the molecular phylogeny as a "clock" with which to examine the timing, and hence potential causes, of evolutionary divergences. However, tentative examination of monophyletic groups within the molecular phylogeny indicates both ancient divergences (e.g. the two *G. ruber* white lineages; Darling *et al.*, 1999) and more recent divergences (possibly within the Quaternary, e.g. *G. bulloides* Type II genotypes).

These difficulties can be addressed in a number of ways:

1. Continue the investigation of genotype distribution within the oceans, since this will hopefully provide additional data with which to support the molecular evidence for cryptic speciation.
2. Investigate using other genes to infer foraminiferal molecular phylogenies. Using other genes should hopefully provide supportive evidence for the topology of the molecular phylogeny based on the SSU rRNA gene.
3. With current molecular phylogenetic techniques the variation in evolution rate within the SSU rRNA gene is too great to establish precise divergence timing. However, it is possible that other genes evolve in a more clock-like fashion and hence would be more suitable for establishing a time calibrated molecular phylogeny.

8.3.2. Application of molecular data to palaeoceanography

If the molecular data, which suggests the presence of cryptic planktic foraminiferal species, is to be incorporated into palaeoceanographic investigations the genotypes/cryptic species must be recognisable. Ideally this would be based on morphology so that this could be applied directly to the marine microfossils. However, there lies a difficulty in establishing a link between genotype and morphotype. Although genotypes within a morphospecies potentially differ in morphology, a genetic divergence may not necessarily be accompanied by a morphological divergence or if morphological divergences do exist between genotypes they may in some cases be very small, as shown by the case of *G. siphonifera* Type I and Type II (Huber *et al.*, 1997). However, the genetic data may provide an indicator as to where further research should concentrate. In particular, this research has highlighted the following morphospecies as candidates for investigation of potential morphological variation: *G. bulloides*, *G. ruber*, *T. quinqueloba*, *G. siphonifera* and *N. pachyderma*. The section below details some experiments that may provide potential breakthroughs.

1. Investigation of genotype/morphotype relationships. This could be addressed by altering the methodology used to collect specimens for DNA analysis. Presently the foraminiferal tests are destroyed during processing and the video imaging was of insufficient quality to determine if morphological subtleties exist between genotypes. If, in addition, to extracting the DNA the foraminiferal test was preserved then the test morphology could be examined by SEM. It is possible that this may be achieved by cell lysis where the foraminiferal cell ruptures due to

osmotic imbalance in solution. The solution containing the DNA could then be processed to preserve it prior to amplification. The test should remain intact for examination. This idea could utilised in culture experiments where genotyping could be incorporated and provide additional biological data.

2. Investigation of test chemistry. If establishing a morphotype/genotype relationship is not feasible then test chemistry may provide further answers. The question of whether genotypes have different chemical signatures due to differing habitat or biological adaptations has enormous implications for palaeoceanographic research based on test chemistry.

How could this be investigated ? Planktic foraminiferal morphospecies represented by multiple genotypes need to be collected from an area known to have a heterogeneous genotype population, such as *G. bulloides* in the North Atlantic current or *G. ruber* in the Canary Island region. Live specimens would be preferable to fossil specimens as genotyping could ensure that the morphospecies in question was represented by multiple genotypes at the time of collection. Improvements in the sensitivity of mass spectroscopy enable analysis of the test chemistry (e.g. $\delta^{18}O$, $\delta^{13}C$, Mg/Ca ratios or Cd/Ca ratios) of individual foraminiferal specimens. By plotting the values graphically (e.g. $\delta^{18}O$ *versus* $\delta^{13}C$) it may be possible to identify different population clusters due to their differing chemical signatures.

3. Single genotype morphospecies. The presence of morphospecies represented by multiple genotypes (e.g. *G. bulloides*) has potential to cause reduced resolution or error for palaeoceanographic investigations. However, *G. falconensis*

was represented by a single genotype in this study. Therefore, it would be of considerable interest to investigate the relative merit of *G. falconensis* test chemistry as an indicator of palaeoceanographic SST's. Indeed, simply by comparing the estimated SST's based on *G. bulloides* and *G. falconensis* test chemistry may show whether it is worth pursuing *G. falconensis* as an additional palaeoceanographic tool.

It is evident that the application of molecular phylogenetics within foraminiferal research is a valued and novel tool. For its full potential to be realised, researchers must strive for ways to apply this data to palaeoceanographic investigations. This section has suggested some ways in which the work presented in this thesis could be extended towards the greater understanding of planktic foraminifers and their application as a palaeoceanographic tool.

References

Anderson, D.M. and Prell, W.L., 1993. A 300 kyr record of upwelling off Oman during the late Quaternary: Evidence of the Asian southwest monsoon. Paleoceanography, 8 (2): 193-208

Bandy, O.L., 1960. The geological significance of coiling ratios in the foraminifer *Globigerina pachyderma* (Ehrenberg). Journal of Paleontology, 34: 671-681.

Bandy, O.L., 1972. Origin and development of *Globorotalia* (*Turborotalia*) *pachyderma* (Ehrenberg). Micropaleontology, 18: 294-318.

Bard, E., Arnold, M, Maurice, P., Duprat, J., Moyes, J. and Duplessy, J-C., 1987. Retreat velocity of the North Atlantic polar front during the last deglaciation determined by 14C mass spectrometry. Nature, 328: 791-794.

Barrett Smith, A., 1963. Distribution of living planktonic foraminifera in the northeast Pacific. Contributions from the Cushman Foundation for Foraminiferal Research, 14 (1): 1-15.

Bauch, H.A., 1994. Significance of variability in *Turborotalita quinqueloba* (Natland) test size and abundance for paleoceanographic interpretations in the Norwegian-Greenland Sea. Marine Geology, 121: 129-141.

Baumann, K.H., Huber, R., Meggers, H. and Henrich, R., 1998. Late Pliocene-early Pleistocene contrasts in carbonate sedimentation between the Norwegian Greenland Sea and the North Atlantic. 6th International Conference on Paleoceanography, Lisbon, p. 76.

Bé, A.W.H., 1960. Ecology of Recent planktonic foraminifera. Part 2- Bathymetric and seasonal distributions in the Sargasso Sea off Bermuda. Micropaleontology, 6 (4): 373-392.

Bé, A.W.H. and Hamlin, W.H., 1967. Ecology of Recent planktonic foraminifera. Part 3- Distribution in the North Atlantic during the summer of 1962. Micropaleontology, 13(1): 87-106.

Bé, A.W.H., 1969. Planktonic foraminifera. Antarctic Map Folio Series, American Geological Society, 11: 9-12.

Bé, A.W.H. and Tolderlund, D.S., 1971. Distribution and ecology of planktonic foraminifera. In: B.M. Funnell and W.R. Riedel, The Micropaleontology of Oceans. Cambridge University Press, London, pp.105-150.

Bé, A.W.H., Harrison, S.M., Frerichs, W.E. and Heiman, M.E., 1976. Variability in test porosity of *Orbulina universa* d'Orbigny at two Indian Ocean localities. In: Y. Takayanagi and T. Saito (Editors), Progress in Micropaleontology. Micropaleontology Press, New York, pp. 1-9.

Bé, A.W.H, Anderson, O.R. and Faber, W.W. Jr., 1983. Sequence of morphological and cytoplasmic changes during gametogenesis in the planktonic foraminifer *Globigerinoides sacculifer* (Brady). Micropaleontology, 29: 310-325

Bé, A.W.H., 1977. An ecological, zoogeographic and taxonomic review of Recent planktonic foraminifera. In: A.T.S. Ramsay, Oceanic Micropaleontology, volume 2, Academic Press, pp. 1-88.

Bemis, B.E., Spero, H.J., Bijma, J. and Lea, D.W., 1998. Reevaluation of the oxygen isotopic composition of planktonic foraminifera: Experimental results and revised paleotemperature equations. Paleoceanography, 13 (2): 150-160.

Berger, W.H., 1970. Planktonic foraminifera: Differential productivity and expatriation off Baja California. Limnology and Oceanography, 15: 183-204.

Berggren, W.A., Kent, D.V., Swisher, C.C. and Aubry, M.P., 1995. A revised Cenozoic geochronology and chronostratigraphy. In: W.A. Berggren, D.V. Kent, and J. Hardenbol (Editors), Geochronology: Time Scales and Global Stratigraphic Correlations. Spec. Publ. Soc. Econ. Paleontol. Mineral., 54: 129-212.

Biermann, C.H., 1998. The molecular evolution of sperm bindin in six species of sea urchin (Echinoida: Strongylocentrotidae). Molecular Biology and Evolution, 15 (2): 1761-1771.

Bijma, J., Faber Jnr, W.W. and Hemleben, C., 1990a. Temperature and salinity limits for growth and survival of some planktonic foraminifers in laboratory cultures. Journal of Foraminiferal Research, 20(2): 95-116.

Bijma, J., Erez, J. and Hemleben, C., 1990b. Lunar and semi-lunar reproductive cycles in some spinose planktonic foraminifers. Journal of Foraminiferal Research, 20(2): 117-127.

Bijma, J., Hemleben, C., Huber, B.T., Erlenkeuser, H. and Kroon, D., 1998. Experimental determination of the ontogenic stable isotope variability in two morphotypes of *Globigerinella siphonifera* (d'Orbigny). Marine Micropaleontology, 35: 141-160.

Blow, W.H., 1969. Late middle Eocene to Recent planktonic foraminiferal biostratigraphy. In: P. Brönnimann and H.H. Renz (Editors), Proceedings of the 1st International Conference on Planktonic Microfossils. Leiden, E.J. Brill, Leiden, Netherlands, volume 1, pp. 199-422.

Blow, W.H., 1979. The Cainozoic Globigerinida: A study of the morphology, taxonomy and evolutionary relationships and the stratigraphical distribution of some Globigerinida. E.J. Brill, Leiden, Netherlands, pp.1413.

Boebel, O., Duncombe Rae, C., Garzoli, S., Lutjeharms, J., Richardson, P., Rossby, T., Schmid, C. and Zenk, W., 1998. Float experiment studies interocean exchanges at the tip of South Africa. EOS, 79 (1): 6-8.

Boltovskoy, E., 1971. Ecology of the planktonic foraminifera living in the surface layer of the Drake Passage. Micropaleontology, 17 (1): 53-68.

Boltovskoy, E., Boltovskoy, D., Correa, N. and Brandini, F., 1996. Planktic foraminifera from the southwestern Atlantic (30°-60°S): species-specific patterns in the upper 50 metres. Marine Micropaleontology, 28: 53-72.

Bradshaw, J.S., 1959. Ecology of living planktonic foraminifera in the north and south equatorial Pacific Ocean. Contributions from the Cushman Foundation for Foraminiferal Research, 10 (2): 25-64.

Brinkhuis, H. and Zachariasse, W. J., 1988. Dinoflagellate cysts, sea level changes and planktonic foraminifers across the Cretaceous-Tertiary boundary at El Haria, Northwest Tunisia. Marine Micropaleontology, 13: 153-191.

Brummer, G.J.A. and Kroon, D., 1988. Genetically controlled planktonic foraminiferal coiling ratios as tracers of past ocean dynamics. In: G.J.A. Brummer and D. Kroon, Planktonic foraminifera as tracers of ocean history. Free University Press, Amsterdam, pp.293-298.

Caron, M. and Homewood, P., 1983. Evolution of early planktic foraminifers. Marine Micropaleontology, 7: 453-462.

Carstens, J. and Wefer, G., 1992. Recent distribution of planktonic foraminifera in the Nansen Basin, Arctic Ocean. Deep-Sea Research, 39 (2):507-524.

Carstens, J., Hebbeln, D. and Wefer, G., 1997. Distribution of planktic foraminifera at the ice margin in the Arctic (Fram Strait). Marine Micropaleontology, 29: 257-269.

Chaisson, W.P. and Pearson, P.N., 1997. Planktonic foraminifer biostratigraphy at Site 925: Middle Miocene-Pleistocene. Proceedings of the Ocean Drilling Program, Scientific Results, 154: 3-32.

Cifelli, R., 1961. *Globigerina incompta*, a new species of pelagic foraminifera from the North Atlantic. Cushman Foundation for Foraminiferal Research, 12 (3): 83-86.

Cifelli, R. and Smith, R.K., 1970. Distribution of planktonic foraminifera in the vicinity of the North Atlantic Current. Smithsonian Contributions to Paleobiology, 4: 1-44.

Cifelli, R., 1971. On the temperature relationships of planktonic foraminifera. Journal of Foraminifera Research, 1 (4): 170-177.

Cifelli, R., 1973. Observations on *Globigerina pachyderma* (Ehrenberg) and *Globigerina incompta* Cifelli from the North Atlantic. Journal of Foraminiferal Research, 3: 157-166.

Cifelli, R. and Bernier, C.S., 1976. Planktonic foraminifera from near the west African coast and a consideration of faunal parcelling in the North Atlantic. Journal of Foraminiferal Research, 6 (4): 258-273.

Clemens, S., Prell, W., Murray, D., Shimmield, G. and Weedon, G., 1991. Forcing mechanisms of the Indian Ocean monsoon. Nature, 353: 720-725.

CLIMAP, 1984. The last interglacial ocean. Quaternary Research, 21: 123-224.

Cordey, WG., 1967. The development of *Globigerinoides ruber* (d'Orbigny 1839) from the Miocene to Recent. Paleontology, 10: 647-659.

Coyne, J.A., 1992. Genetics and speciation. Nature, 355: 511-515.

Coyne, J.A. and Orr, H.A., 1998. The evolutionary genetics of speciation. In: Evolution of biological diversity: from population differentiation to speciation, Philosophical Transactions of the Royal Society, Series B, 355 (1366): 287-305.

Cronin, T.M. and Dowsett, H.J., 1996. Biotic and oceanographic response to the

Pliocene closing of the central American Isthmus. In: J.B.C. Jackson, A.F. Budd and A.G. Coates (Editors). Evolution and environment in tropical America, University of Chicago Press, Chicago, pp. 76-104.

Culver, S.J., 1991. Early Cambrian Foraminifera from West Africa. Science, 254: 689-691.

Darling, K.F., Kroon, D., Wade, C.M. and Leigh Brown, A.J., 1996a. The isolation and amplification of the 18s ribosomal RNA gene from planktonic foraminifers using gametogenic specimens. In: A. Moguilevsky and R.C. Whatley (Editors), Microfossils and Oceanic Environments. Proceedings of 'ODP and the Marine Biosphere' International Conference. Aberystwyth Press, Chapter 3.1, pp. 249-259.

Darling, K.F., Kroon, D., Wade, C.M. and Leigh Brown, A.J., 1996b. Molecular phylogeny of planktic foraminifera. Journal of Foraminiferal Research, 26: 324-330.

Darling, K.F., Wade, C.M., Kroon, D. and Leigh Brown, A.J., 1997. Planktic foraminiferal molecular evolution and their polyphyletic origins from the benthic taxa. Marine Micropaleontology, 30: 251-266.

Darling, K.F., Wade, C.M., Kroon, D. and Leigh Brown, A.J., 1999. The diversity and distribution of modern planktic foraminiferal SSU rRNA genotypes and their potential as tracers of present and past ocean circulations. Paleoceanography, 14 (3): 3-12.

Darling, K.F., Wade, C.M., Stewart, I.A., Kroon, D., Dingle, R. and Leigh Brown, A.J., 2000. Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers. Nature, 405: 43-47.

Deitrich, G., 1969. Atlas of the hydrography of the northern North Atlantic Ocean. Cons. Int. Explot. Mer., 140 pp.

De Rijk, P., Van de Peer, Y. and De Wachter, R., 1996. Database on the structure of large ribosomal subunit RNA. Oxford University Press, pp. 92-97.

de Vargas, C., Zaninetti, L., Hilbrecht, H. and Pawlowski, J., 1997. Phylogeny and rates of molecular evolution of planktonic foraminifera: SSU rDNA sequences compared to the fossil record. Journal of Molecular Evolution, 45 (3): 285-294.

de Vargas, C. and Pawlowski, J., 1998. Molecular versus taxonomic rates of

evolution in planktonic foraminifera. Molecular Phylogenetics and Evolution, 9: 463-469.

de Vargas, C., Norris, R., Zaninetti, L., Gibb, S. W. and Pawlowski, J., 1999. Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. Proceedings of the National Academy of Science, 96: 2864-2868.

Deuser, W.G., Ross, E.H., Hemleben, C. and Spindler, M., 1981. Seasonal changes in species composition, numbers, mass, size, and isotopic composition of planktonic foraminifera settling into the deep Sargasso Sea. Palaeogeography, Palaeoclimatology, Palaeoecology, 33: 103-127.

Deuser, W.G., 1987. Seasonal variations in isotopic composition and deep-water fluxes of the tests of perennially abundant planktonic foraminifera of the Sargasso Sea: results from sediment-trap collections and their paleoceanographic significance. Journal of Foraminiferal Research, 17 (1): 14-27.

Deuser, W.G. and Ross, E.H., 1989. Seasonally abundant planktonic foraminifera of the Sargasso Sea: succession, deep-water fluxes, isotopic compositions, and paleoceanographic implications. Journal of Foraminiferal Research, 19(4): 268-293.

Dobzhansky, T. 1937. *Genetics and the origin of the species*. Columbia University Press, New York.

Duque-Caro, H., 1990. Neogene stratigraphy, paleoceanography, and paleobiology in northwest South America and the evolution of the Panama Seaway. Paleogeography, Paleoclimatology, Paleoecology, 77: 203-234.

Elderfield, H., Rickaby, R., Chapman, M. and Greaves, M., 1998. Reconstructing past sea temperatures and nutrients using Mg/Ca and Cd/Ca in planktonic foraminifera. 6th International Conference on Paleoceanography, Lisbon, p.102.

Epstein, S., Buchsbaum, R., Lowenstaum, H.A. and Urey, H.C., 1953. Revised carbonate-water isotopic temperature scale. Geological Society of America Bulletin, 64: 1315-1325.

Ericson, D.B., Wollin, G. and Wollin, J., 1954. Coiling direction of *Globorotalia truncatulinoides* in deep-sea cores. Deep-Sea Research, 2: 152-158.

Ericson, D.B. and Wollin, G., 1956. Correlation of six cores from the equatorial Atlantic and the Caribbean. Deep-Sea Research, 3: 104-125.

Ericson, D.B., 1959. Coiling direction of *Globigerina pachyderma* as a climatic index. Science, 130: 219-220.

Faber Jr, W.W., Anderson, O.R., Lindsey, J.L. and Caron, D.A., 1988. Algal-foraminiferal symbiosis in the planktonic foraminifer *Globigerinella aequilateralis*: i. Occurrence and stability of two mutually exclusive chrysophyte endosymbionts and their ultrastructure. Journal of Foraminiferal Research, 18 (4): 334-343.

Faber Jr, W.W., Anderson, O.R., and Caron, D.A., 1989. Algal-foraminiferal symbiosis in the planktonic foraminifer *Globigerinella aequilateralis*: ii. Effects of two symbiont species on foraminiferal growth and longevity. Journal of Foraminiferal Research, 19: 185-193.

Fairbanks, R.G., Weibe, P.H. and Bé, A.W.H., 1980. Vertical distribution and isotopic composition of living planktonic foraminifera in the western North Atlantic. Science, 207: 61-63.

Fairbanks, R.G., Sverdlove, M., Free, R., Wiebe, P.H. and Bé, A.W.H., 1982. Vertical distribution and isotopic fractionation of living planktonic foraminifera from the Panama Basin. Nature, 298: 841-844.

Felsenstein, J., 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution, 39: 783-791.

Felsenstein, J., 1993. PHYLIP, Manual version 3.52c. Berkeley University Herbarium, University of California, Berkeley, California.

Fitch, W.M. and Margoliash, E., 1967. Construction of phylogenetic trees: A method based on mutation distances as estimated from cytochrome c sequences is of general applicability. Science, 155: 279-284.

Funnell, B.M. and Swallow, J.E., 1997. Intra-sample, inter-sample and down-core microvariation in sea-surface temperature estimates obtained from planktonic foraminifera in the NE Atlantic. Journal of Micropalaeontology, 16; 163-174.

Ganssen, G.M. and Kroon, D., in press. The isotopic signature of planktic foraminifera from Northeast Atlantic surface sediments: implications for the reconstruction of past oceanic conditions. Journal of the Geological Society.

Gast, R.J. and Caron, D.A., 1996. Molecular phylogeny of symbiotic dinoflagellates from planktonic foraminifera and radiolaria. Molecular Biology and Evolution, 13 (9): 1192-1197.

Giraudeau, J., 1993. Planktonic foraminiferal assemblages in surface sediments from the southwest African continental margin. Marine Geology, 110: 47-62.

Görög, A., 1994. Early Jurassic planktonic foraminifera from Hungary. Micropaleontology, 40 (3): 255-260.

Gower, J.F.R., Denman, K.L. and Holyer, R.J., 1980. Phytoplankton patchiness indicates the fluctuation spectrum of mesoscale oceanic structure. Nature, 288: 157-159.

Haynes, J.R. 1992. Supposed pronounced ecophenotypy in foraminifera. Journal of Micropalaeontology, 11 (1): 59-63.

Hecht, A.D., Bé, A.W.H. and Lott, L., 1976. Ecologic and paleoclimatic implications of morphological variation of *Orbulina universa* in the Indian Ocean. Science, 194: 422-424.

Hemleben, C., Spindler, M., Breitinger, I. And Ott, R., 1987. Morphological and physiological responses of *Globigerinoides sacculifer* (Brady) under varying laboratory conditions. Marine Micropaleontology, 12: 305-324.

Hemleben, C., Spindler, M. and Anderson, O.R., 1989. Modern planktonic foraminifera. Springer, New York, NY, 335pp.

Hilbrecht, H., 1997. Morphological gradation and ecology in *Neogloboquadrina pachyderma* and *N. dutertrei* (planktic foraminifera) from core top sediments. Marine Micropaleontology, 31: 31-43.

Hillis, D.M., Larson, A., Davis, S.K. and Zimmer, E.A., 1990. Nucleic Acids III: Sequencing. In: D.M. Hillis and C. Moritz (Editors), Molecular Systematics. Sinauer Associates, Sunderland, Massachusetts, pp.318-372.

Hillis, D.M. and Dixon, M.T., 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Quarterly Review Biology, 66 (4): 411-446.

Hillis, D.M. and Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biology, 42: 182-192.

Hillis, D.M., Moritz, C. and Mable, B.K. (Editors), 1996. Molecular Systematics, 2nd edition. Sinauer Associates Inc., Sunderland, Massachusetts, USA.

Holzmann, M., Piller, W. and Pawlowski, J., 1996. Sequence variations in the large-subunit ribosomal RNA gene of *Ammonia* (Foraminifera, Protozoa) and their evolutionary implications. Journal of Molecular Evolution, 43: 145-151.

Holzmann, M. and Pawlowski, J., 1996. Preservation of foraminifera for DNA extraction and PCR amplification. Journal of Foraminiferal Research, 26: 264-267.

Huber, B.T., Bijma, J. and Darling, K., 1997. Cryptic Speciation in the living planktonic foraminifer *Globigerinella siphonifera* (d'Orbigny). Paleobiology, 23 (1): 33-62.

Huber, R., Meggers, H., Baumann, K.H., Raymo, M.E. and Henrich, R., 1998. Morphometric evolution of *Neogloboquadrina pachyderma* sin. in the Norwegian-Greenland Sea during the last 1.3 Ma – Implications for paleoceanographic reconstructions. 6th International Conference on Paleoceanography, Lisbon, p.129.

Imbrie, J. and Kipp, J.Z., 1971. A new micropaleontological method for quantitative paleoclimatology: Application to a late Pleistocene Caribbean core. In: K.K. Turekian (Editor), The late Cenozoic glacial ages, Yale University Press, U.S.A., pp. 77-181.

Ivanova, E.M., Conan, S.M.H., Peeters, J.F.C. and Troelstra, S.R., 1999. Living *Neogloboquadrina* sin and its distribution in the sediments from Oman and Somalia upwelling areas. Marine Micropaleontology, 36: 91-107.

Keigwen Jr, L.D., 1978. Pliocene closing of the Isthmus of Panama, based on biostratigraphic evidence from nearby Pacific Ocean and Caribbean Sea cores. Geology, 6: 630-634.

Keigwen, L., 1982. Isotopic paleoceanography of the Caribbean and east Pacific: Role of Panama uplift in late Neogene time. Science, 217: 350-353.

Kellog, T.B., 1984. Late-Glacial - Holocene high-frequency climatic changes in deep-sea cores from the Denmark Strait. In: A. Mörner and W. Karlén, Climatic Changes on a Yearly to Millennial Basis, D. Reidel Publishing Company, pp. 123-132.

Kennett, J.P., 1968. Latitudinal variation in *Globigerina pachyderma* (Ehrenberg) in surface sediments of the southwest Pacific Ocean. Micropaleontology, 14: 305-318.

Kennett, J.P., 1969. Distribution of planktonic foraminifera in surface sediments to the southeast of New Zealand. IN: P. Brönnimann and H.H. Renz (Editors), Proceedings of the 1st International Conference on Planktonic Microfossils. Leiden, E.J. Brill, Leiden, Netherlands, volume 2, pp. 307-322.

Kennett, J.P. and Srinivasan, M.S., 1980. Surface ultrastructure variation in *Neogloboquadrina pachyderma* (Ehrenberg): phenotypic variation and phylogeny in the late cenozoic. Cushman Foundation Special Publication, 19: 134-162.

Kennett, J.P. and Srinivasan, M.S., 1983. Neogene Planktonic Foraminifera, A Phylogenetic Atlas. Hutchinson and Ross, Stroudsburg, PA, 263pp.

King Jr, K. and Hare, P.E., 1972a. Amino acid composition of planktonic foraminifera: paleobiochemical approach to evolution. Science, 175: 1461-1463.

King Jr, K. and Hare, P.E., 1972b. Amino acid composition of the test as a taxonomic character for living and fossil planktonic foraminifera. Micropaleontology, 18 (3): 285-293.

Kishino, H. and Hasegawa, M., 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data and the branching order of the Hominoidea. Journal of Molecular Evolution, 4: 406-425.

Knoll, A.H., 1992. The early evolution of the eukaryotes: A geological perspective. Science, 256: 622-627.

Knowlton, N., 1993. Sibling species in the sea. Annual Review of Ecolological Systematics, 24: 189-216.

Knowlton, N., Weigt, L.A., Solórzano, L.A., Mills, D.K. and Bermingham, E., 1993. Divergence in proteins, mitochondrial DNA, and reproductive compatibility across the Isthmus of Panama. Science, 260: 1629-1632.

Kroon, D., Wouters, P.F., Moodley, L., Ganssen, G. and Troelstra, S.R., 1988. Phenotypic variation of *Turborotalita quinqueloba* (Natland) tests in living populations and in the Pleistocene of an eastern Mediterranean piston core. In: G.J.A. Brummer and D. Kroon, Planktonic foraminifera as tracers of ocean history. Free University Press, Amsterdam, pp. 131-147.

Kroon, D. and Ganssen, G., 1989. Northern Indian Ocean upwelling cells and the stable isotope composition of living planktonic foraminifers. Deep-Sea Research, 36 (8): 1219-1236.

Kroon, D., 1991. Distribution of extant planktic foraminiferal assemblages in Red Sea and northern Indian Ocean surface waters. Revista Española de Micropaleontologia, 23 (1): 37-74.

Kroon, D. and Darling, K., 1995. Size and upwelling control of the stable isotope composition of *Neogloboquadrina dutertrei* (d'Orbigny), *Globigerinoides ruber* (d'Orbigny) and *Globigerina bulloides* d'Orbigny: Examples from the Panama Basin and Arabian Sea. Journal of Foraminiferal Research, 25 (1): 39-52.

Kroon, D., Austin, W. E. N., Chapman, M. R. and Ganssen, G. M., 1997. Deglacial surface circulation changes in the northeastern Atlantic: Temperature and salinity records off NW Scotland on a century scale. Paleoceanography, 12 (6):755-763.

Langer, M.R., Lipps, J.H. and Piller, W.E., 1993. Molecular paleobiology of protists: amplification and direct sequencing of foraminiferal DNA. Micropaleontology, 39: 63-68.

Lazarus, D., 1983. Speciation in pelagic Protista and its study in the planktonic microfossil record: a review. Paleobiology, 9 (4): 327-340.

Lazarus, D., Hilbrecht, H., Spencer-Cervato, C. and Thierstein, H., 1995. Sympatric speciation and phyletic change in *Globorotalia truncatulinoides*. Paleobiology, 21 (1): 28-51.

Lessios, H.A., 1984. Possible prezygotic reproductive isolation in sea urchins separated by the Isthmus of Panama. Evolution, 38 (5): 1144-1148.

Lessios, H.A. and Cunningham, C.W., 1990. Gametic incompatibility between species of the sea urchin *Echinometra* on the two sides of the Isthmus of Panama. Evolution, 44 (4): 933-941.

Levitus, S., Burgett, R. and Boyer, T.P., 1994. World Ocean Atlas volume 3: Salinity. NOAA Atlas NESDIS 3, World Ocean Atlas 1994, U.S. Department of Commerce, Washington D.C., 99 pp.

Levitus, S. and Boyer, T.P., 1994. World Ocean Atlas 1994, volume 4: Temperature. NOAA Atlas NESDIS 4, U.S. Department of Commerce, Washington D.C., 117 pp.

Lindberg, D.R., 1991. Marine biotic interchange between the northern and southern hemispheres. Paleobiology, 17 (3): 308-324.

Lipps, J., 1979. Ecology and paleoecology of planktic foraminifera. In: J.H. Lipps, W.H. Berger, M.A. Buzas, R.G. Douglas and C.A. Ross (Editors), Foraminiferal Ecology and Paleoecology, SEPM Short Course No. 6, Houston, pp. 62-104.

Little, M.G., Schneider, R.R., Kroon, D., Price, B., Bickert, T. and Wefer, G., 1997. Rapid palaeoceanographic changes in the Benguela Upwelling System for the last 160,000 years as indicated by abundances of planktonic foraminifera. Palaeogeography, Palaeoclimatology, Palaeoecology, 130: 135-161.

Loeblich, Jr., A.R. and Tappan, H., 1974. Recent advances in the classification of Foraminiferida. In: R.H. Hedley and C.G. Adams (Editors), Foraminifera, 1. Academic Press, London, PP. 1-53.

Lohmann, G.P., 1992. Increased seasonal upwelling in the subtropical South Atlantic over the last 700, 000 years: evidence from deep-living planktonic foraminifera. Marine Micropaleontology, 19: 1-12.

Lutjeharms, J.R.E., 1996. The exchange of water between the South Indian and South Atlantic Oceans. In: G. Wefer, W.H. Berger, G. Siedler and D. Webb, The South Atlantic: Past and Present Circulation, Springer-Verlag, Berlin, pp. 122-162.

Maidak, B.L., Larsen, N., McCaughey, M.J., Overbeek, R., Olsen, G.J, Fogel, K., Blandy, J. and Woese, C.R., 1994. The ribosomal database project. Nucleic Acids Research, 22: 3484-3487.

Maidak, B.L., Olsen, G.J, Larsen, N., Overbeek, R., McCaughey, M.J. and Woese, C.R., 1996. The Ribosomal Database Project (RDP), Oxford University Press, Oxford, pp. 82-85.

Malmgren, B. and Kennett, J.P., 1972. Biometric analysis of phenotypic variation: *Globigerina pachyderma* (Ehrenberg) in the South Pacific Ocean. Micropaleontology, 18 (2): 241-248.

Malmgren, B. and Kennett, J.P., 1976. Biometric analysis of phenotypic variation in Recent *Globigerina bulloides* d'Orbigny in the southern Indian Ocean. Marine Micropaleontology, 1: 3-25.

Malmgren, B. and Kennett, J.P., 1977. Biometric differentiation between Recent *Globigerina bulloides* and *Globigerina* falconensis in the Southern Indian Ocean. Journal of Foraminiferal Research, 7 (2): 130-148.

Malmgren, B. and Kennett, J.P., 1978. Test size variation in *Globigerina bulloides* in response to Quaternary palaeoceanographic changes. Nature, 275: 123-124.

Marchitto Jr., T.M., Curry, W.B. and Oppo, D.W., 1998. Millenial-scale changes in North Atlantic circulation since the last glaciation. Nature, 393: 557-561.

Maynard Smith, J., 1966. Sympatric speciation. The American Naturalist, 100 (916): 637-650.

Mayr, E., 1942. Systematics and the Origin of Species, Columbia University Press, New York.

McIntyre, A., Ruddiman, W.F., Karlin, K. and Mix, A.C., 1989. Surface water response of the equatorial Atlantic Ocean to orbital forcing. Paleoceanography, 4 (1): 19-55.

Meggers, H., Feudenthal, T., Davenport, B., Neuer, S. and Wefer, G., 1998. Particle flux and paleoceanography in the eastern boundary current system. 6th International Conference on Paleoceanography, Lisbon, p. 164.

Merle, C., Moullarde, M., Lima, O. and Perasso, R., 1994. Essai de caractérisation phylogénétique de Foraminifères planctoniques a partir de séquences partielles d'ADNr28s. Paleontology (Evolution, Micropaleontology), 319 (2):149-153)

Metz, E.C. and Palumbi, S.R., 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. Molecular Biology and Evolution, 13 (2): 397-406.

Mix, A.C. and Morey, A.E., 1996. Climate feedback and Pleistocene variations in the Atlantic south equatorial current. In: G. Wefer, W.H. Berger, G. Siedler and D. Webb, The South Atlantic: Past and Present Circulation, Springer-Verlag, Berlin, pp. 503-525.

Mix, A.C., Morey, A.E., Pisias, N.G. and Hostetler, S.W., 1999. Foraminiferal faunal estimates of paleotemperature: circumventing the no-analog problem yields cool ice-age tropics. Paleoceanography, 14 (3): 350-359.

Mullis, K.B. and Faloona, F.A., 1987. Specific synthesis of DNA *in vitro* via a polymerase-catalysed chain reaction. Methods in Enzymology, 155: 335-350.

Naidu, P.D. and Malmgren, B.A., 1996a. Relationship between Late Quaternary upwelling history and coiling properties of *Neogloboquadrina pachyderma* and *Globigerina bulloides* in the Arabian Sea. Journal of Foraminiferal Research, 26 (1):64-70.

Naidu, P.D. and Malmgren, B.A., 1996b. A high resolution record of late Quaternary upwelling along the Oman Margin, Arabian Sea, based on planktonic foraminifera. Paleoceanography, 11 (1):129-140.

Neefs, J.M., Van de Peer, Y., De Rijk, P., Chapelle, S. and De Wachter, R., 1993. Compilation of small ribosomal subunit RNA structures. Nucleic Acids Research, 21 (13): 3025-3049.

Neibler, H.S. and Gersonde R., 1998. A planktic foraminiferal transfer function for the southern South Atlantic Ocean. Marine Micropaleontology, 34: 213-234.

Norris, R.D., 1991. Biased extinction and evolutionary trends. Paleobiology, 17(4): 388-399.

Norris, R.D., Corfield, R.M. and Cartlidge, J., 1996. What is gradualism ? Cryptic speciation in globorotaliid foraminifera. Paleobiology, 22 (3): 386-405.

Norris, R.D., 1999. Hydrographic and tectonic control of plankton distribution and evolution. In: F. Abrantes and A.C. Mix (Editors), Proceedings for the 6th ICP, 1998, Reconstructing Ocean History: A Window into the Future, Kluwer Academic/Plenum Publishers, New York, pp. 173-193.

Olsen, G.J. and Woese, C.R., 1993. Ribosomal RNA: a key to phylogeny. Federation of American Societies for Experimental Biology, 7: 113-123.

Olsson, R.K., Hemleben, C., Berggren, W.A. and Liu, C., 1992. Wall texture classification of planktonic foraminifera genera in the lower Danian. Journal of Foraminiferal Research, 22(3): 195-213.

Olsson, R.K., 1976. Wall structure, topology, and crust of *Globigerina pachyderma* (Ehrenberg). In: Y. Takayanagi and T. Saito (Editors), Progress in Micropaleontology. Micropaleontology Press, New York, pp. 244-257.

Ortiz, J.D., Mix, A.C. and Collier, R.W., 1995. Environmental control of living symbiotic and asymbiotic foraminifera of the California Current. Paleoceanography, 10 (6): 987-1009.

Ottens, J.J., 1991. Planktic foraminifera as North Atlantic water mass indicators.

Oceanol. Acta., 14 (2): 123-140.

Ottens, J.J. (Editor), 1992. Planktic foraminifera as indicators of oceanic environments in the Northeast Atlantic, Free University Press, Amsterdam, Netherlands, pp. 189.

Page, R.D.M. and Holmes, E.C. (Editors), 1998. Molecular Evolution, A Phylogenetic Approach, Blackwell Science Ltd, Oxford.

Palumbi, S.R., 1994. Genetic divergence, reproductive isolation, and marine speciation. Annual Review of Ecological Systematics, 25: 574-572.

Parker, F.L., 1962. Planktonic foraminiferal species in Pacific sediments. Micropaleontology, 8 (2): 219-254.

Parker, F.L. and Berger, W.H., 1971. Faunal and solution patterns of planktonic foraminifera in surface sediments of the South Pacific. Deep-Sea Research, 18: 73-107.

Pawlowski, J., Bolivar, I., Fahrni, J. and Zaninetti, L., 1994a. Taxonomic identification of foraminifera using ribosomal DNA sequences. Micropaleontology, 40: 373-377.

Pawlowski, J., Bolivar, I., Guiard-Maffia, J. and Gouy, M., 1994b. Phylogenetic position of foraminifera inferred from LSU rRNA gene sequences. Molecular Biology and Evolution, 11(6): 929-938.

Pawlowski, J., Bolivar, I., Fahrni, J., Cavalier-Smith, T. and Gouy, M., 1996. Early evolution of foraminifera suggested by SSU rRNA gene sequences. Molecular Biology and Evolution, 13 (3): 445-450.

Pawlowski, J., Bolivar, I., Fahrni, J. and de Vargas, C., 1997. Extreme differences in rates of molecular evolution of foraminifera revealed by comparison of ribosomal DNA sequences and the fossil record. Molecular Biology and Evolution, 14: 498-505.

Pawlowski, J., Bolivar, I., Fahrni, J., de Vargas, C. and Bowser, S.S., 1999. Naked foraminiferans revealed. Nature, 399: 27.

Pearson, P.N., 1993. A lineage phylogeny for the Paleogene planktonic foraminifera. Micropaleontology, 39: 193-232.

Pflaumann, U., Duprat, J., Pujol, C. and Labeyrie, L.D., 1996. SIMMAX: A modern analog technique to deduce Atlantic sea surface temperatures from planktonic foraminifera in deep-sea sediments. Paleoceanography, 11 (1): 15-35.

Pujol, C. and Grazzini, C.V., 1995. Distribution patterns of live planktic foraminifers as related to regional hydrography and productive systems of the Mediterranean Sea. Marine Micropaleontology, 25: 187-217.

Reynolds, L.A. and Thunell, R.C., 1986. Seasonal production and morphologic variation of *Neogloboquadrina pachyderma* (Ehrenberg) in the northeast Pacific. Micropaleontology, 32 (1): 1-18.

Rickaby, R. and Elderfield, H., 1999. Planktonic foraminiferal Cd/Ca: paleonutrients or paleotemperature ?. Paleoceanography, 14 (3): 293-303.

Robbins, L.L. and Healy-Williams, N., 1991. Toward a classification of planktonic foraminifera based on biochemical, geochemical and morphological criteria. Journal of Foraminiferal Research, 21 (2): 159-167.

Robbins, L.L., 1988. Environmental significance of morphotypic variability in open-ocean versus ocean-margin assemblages of *Orbulina universa*. Journal of Foraminiferal Research, 18 (2):326-333.

Saiki, R.K., Gelfland, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Mullis, K.B. and Ehrlich, H.A., 1988. Primer-directed enzymatic amplification of DNA with thermostatic DNA polymerase. Science, 239: 487-490.

Saito, T., 1976, Geologic significance of coiling direction in the planktonic foraminifera *Pulleniatina*. Geology, 4: 305-309.

Saito, T., Thompson, P.R. and Berger, D., 1981. Systematic index of Recent and Pleistocene planktonic foraminifera. University of Tokyo Press, Japan, pp. 190.

Saitou, N. and Nei, M., 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 4: 406-425.

Sautter, L.R. and Thunell, R.C., 1991a. Planktonic foraminiferal response to upwelling and seasonal hydrographic conditions: sediment trap results from San Pedro Basin, Southern California Bight. Journal of Foraminiferal Research, 21 (4): 347-363.

Sautter, L.R. and Thunell, R.C., 1991b. Seasonal variability in the $\delta^{18}O$ and $\delta^{13}C$ of planktonic foraminifera from an upwelling environment: sediment trap results

from the San Pedro Basin, Southern California Bight. Paleoceanography, 6 (3): 307-334.

Schiebel, R., Bijma, J. and Hemleben, C., 1997. Population dynamics of the planktic foraminifer *Globigerina bulloides* from the eastern North Atlantic. Deep-Sea Research I, 44 (9-10):1701-1713.

Schiebel, R., 1999. Foraminifers and pteropods. In: Poseidon 247 cruise report, Institut für Meeresskunde an der Universität Kiel, Germany, p. 9.

Smith, S.W., Overbeek, R., Woese, C.R., Gilbert, W. and Gilevet, P.M., 1994. The genetic data environment and expandable GUI for multiple sequence analysis. Comp. Appl. Biosci., 10: 671-675.

Sogin, M.L., 1989. Evolution of eukaryotic microorganisms and their small subunit ribosomal RNAs. Am. Zool, 29: 487-499.

Sogin, M.L., 1991. Early evolution and the origin of the eukaryotes. Curr. Opin. Genet. Dev., 1: 457-463.

Spero, H.J. and Lea, D.W., 1996. Experimental determination of stable isotope variability in *Globigerina bulloides*: implications for paleoceanographic reconstructions. Marine Micropaleontology, 28: 231-246.

Spezzaferi, S. and Premoli Silva, I., 1991. Oligocene planktonic biostratigraphy and paleoclimate interpretation from hole 538A, DSDP Leg 77, Gulf of Mexico. Palaeogeography, Palaeoclimatology and Palaeoecology, 83: 217-263.

Spezzaferri, S., 1994. Planktonic foraminiferal biostratigraphy and taxonomy of the Oligocene and lower Miocene in the oceanic record: An overview. Paleontographia Italica, 81: 1-187.

Spindler, M., Anderson, O.R., Hemleben, C. and Bé, A.W.H., 1978. Light and electron microscopic observations in gametogenesis in *Hastigerina pelagica* (Foraminifera). Journal of Protozoology, 25: 427-433.

Swift, J.H., 1986. The Arctic Waters. In B.G. Hurdle (Editor), The Nordic Seas. Springer, New York, pp. 129-151.

Swofford, D.L. and Olsen, G.J., 1990. Phylogenetic Reconstruction. In: D.M. Hillis and C. Moritz (Editors), Molecular Systematics. Sinauer Associates, Sunderland, Massachusetts, pp.411-501.

Swofford, D.L., Olsen, G.L., Waddell, P.J. and Hillis, D.M., 1996. Phylogenetic Inference. In: D.M. Hillis, C. Moritz and B.K. Mable (Editors), Molecular Systematics. 2nd Edition, Sinauer Associates, Sunderland, Massachusetts, pp.407-425.

Tappan, H. and Loeblich, A.R., 1988. Foraminiferal evolution, diversification, and extinction. Journal of Paleontology, 62(5): 695-714.

Thiede, J., 1975. Distribution of foraminifera in surface waters of a coastal upwelling area. Nature, 253: 712-714.

Thompson, P.R., Duplessy, A.W.H. and Shackelton, N.J., 1969. Disappearance of pink-pigmented *Globigerinoides ruber* at 120,000 years BP in the Indian and Pacific Oceans. Nature, 280: 554-558.

Tolderlund, D.S. and Bé, A.W.H., 1971. Seasonal distribution of planktonic foraminifera in the western North Atlantic. Micropaleontology, 17 (3): 297-329.

Ufkes, E. and Zachariasse, W., 1993. Origin of coiling differences in living *Neogloboquadrinids* in the Walvis Bay region, off Namibia, southwest Africa. Micropaleontology 39 (3):283-287.

Ufkes, E., Jansen, J.H.F. and Brummer, G.J.A., Living planktonic foraminifera in the eastern South Atlantic during spring: indicators of water masses, upwelling and the Congo (Zaire) River plume. Marine Micropaleontology, 33: 27-53.

Vermeij, G.J., 1991. Anatomy of an invasion: the trans-Arctic interchange. Paleobiology, 17 (3): 281-307.

Wade, C.M., Darling, K.F., Kroon, D. and Leigh Brown, A.J., 1996. Early evolutionary origin of planktic foraminifera inferred from small subunit rDNA sequence comparisons. Journal of Molecular Evolution, 43: 672-677.

Wainwright, P.O., Hinkle, G., Sogin, M.L. and Stickel, S.K., 1993. Monophyletic origins of the metazoa: an evolutionary link with fungi. Science, 260: 340-342.

Washburn, L., Emery, B.M., Jones, B.H. and Ondercin, D.G., 1998. Eddy stirring and phytoplankton patchiness in the subarctic North Atlantic in late summer. Deep-Sea Research, 45: 1411-1439.

White, T.J., Bruns, T., Lee, S. and Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: M.A. Innis, D.H. Gelfland, J.J. Sninsky and T.J. White (Editors), PCR Protocols: A Guide to

Methods and Applications. Harcourt Brace Jovanovich, San Diego, California, pp. 315-322.

Wray, C.G., Lee, J.J. and DeSalle, R., 1993. Extraction and enzymatic characterisation of foraminiferal DNA. Micropaleontology, 39(1): 69-73.

Zhao, M., Beveridge, N.A.S., Shackleton, N.J., Sarnthein, M. and Eglington, G., 1995. Molecular stratigraphy of cores off northwest Africa: sea surface temperature history over the last 80 ka. Paleoceanography, 10 (3), 661-675.

Appendices Index

A1.1. Abbreviations used in the alignment and distance matrix for the 505 bp molecular phylogeny

C_G.sI	*G. siphonifera* Type I (Caribbean)
NA_sI	*G. siphonifera* Type I (North Atlantic)
C_G.sI(V)	*G. siphonifera* Type I (Caribbean, de Vargas)
C_G.sIIa	*G. siphonifera* Type IIa (Caribbean)
CS_G.sIIa	*G. siphonifera* Type IIa (Coral sea)
NA_G.sIIa	*G. siphonifera* Type IIa (North Atlantic)
CA_G.sIIa	*G. siphonifera* Type IIa (S. Californian Bight)
CA_G.sIIb	*G. siphonifera* Type IIb (S. Californian Bight)
NA_G.sIIb	*G. siphonifera* Type IIb (North Atlantic)
NA_cal	*G. calida* (North Atlantic)
C_O.uni	*O. universa* (Caribbean)
CS_O.uni	*O. universa* (Coral Sea)
CA_O.uni	*O. universa* (S. Californian Bight)
M_O.uni(V)	*O. universa* (Mediterranean, de Vargas)
C_G.sac	*G. sacculifer* (Caribbean)
CS_G.sac	*G. sacculifer* (Coral Sea)
C_G.rub,P	*G. ruber* pink (Caribbean)
NA_G.rub,P	*G. ruber* pink (North Atlantic)
NA_G.rub,WI	*G. ruber* white Type I (North Atlantic)
CS_G.rub,WI	*G. ruber* white Type I (Coral Sea)
C_G.rub(P)	*G. ruber* (Caribbean, Pawlowski)
NA_G.rub,WII	*G. ruber* white Type II (North Atlantic)
CA_G.rub,WII	*G. ruber* white Type II (S. Californian Bight)
C_G.con(P)	*G. conglobatus* (Caribbean, Pawlowski)
CS_G.con	*G. conglobatus* (Coral Sea)
NA_N.pac(d)	*N. pachyderma*, dextral (North Atlantic)
AA_N.pac(d)	*N. pachyderma*, dextral (subantarctic)
C_N.dut	*N. dutertrei* (Caribbean)
CS_G.bul,Ia	*G. bulloides* Type Ia (Coral Sea)
CA_G.bul,IId	*G. bulloides* Type IId (S. Californian Bight)
M_G.bul(V)	*G. bulloides* Type Ib (Mediterranean)
NA_G.bul,Ib	*G. bulloides* Type Ib (North Atlantic)
NA_G.bul,IIb	*G. bulloides* Type IIb (North Atlantic)
AA_G.bul,IIb	*G. bulloides* Type IIb (subantarctic)
NA_G.bul,IIa	*G. bulloides* Type IIa (North Atlantic)
AA_G.bul,IIa	*G. bulloides* Type IIa (subantarctic)
AA_G.bul,IIc	*G. bulloides* Type IIc (subantarctic)
NA_G.falc	*G. falconensis* (North Atlantic)
CS_G.falc	*G. falconensis* (Coral Sea)
NA_T.qui,IIa	*T. quinqueloba* Type IIa (North Atlantic)
AA_T.qui,IIa	*T. quinqueloba* Type IIa (subantarctic)
NA_T.qui,IIb	*T. quinqueloba* Type IIb (North Atlantic)
AA_T.qui,IIc	*T. quinqueloba* Type IIc (subantarctic)
CS_T.qui,I	*T. quinqueloba* Type I (Coral Sea)
CS_G.glu	*G. glutinata* (Coral Sea)
NA_G.uvu	*G. uvula* (North Atlantic)

A.becca_P	*Ammonia beccarii*
Troch_P	*Trochammina hadai*
H.germ_P	*Haynesina germanica*
Textu_P	*Textularia* sp.
Biger_P	*Bigerina* sp.
Boliv_P	*Bolivina* sp.
G.oper_P	*Glabratella opercularis*
A.trian_P	*Astrorhiza triangularis*
As.rara_P	*Astrammina rara*
A.angu_P	*Archaias angulatus*
E.acul_P	*Elphidium aculeatum*
Quin_P	*Quinqueloculina* sp
M.secan_P	*Massilina secans*
P.pert_P	*Peneroplis pertusus*
Allogrom_P	*Allogromia* sp.

A1.2. Alignment for 505 bp molecular phylogeny

```
  C_G.sI        1  GCA--CCACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
 NA_G.sI        1  GCA--CCACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
C_G.sI(V)       1  GCA--CCACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
C_G.sIIa        1  GCA--CTACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
NA_G.sIIa       1  GCA--CCACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
CS_G.sIIa       1  GCA--CTACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
CA_G.sIIa       1  GCA--CTACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
CA_G.sIIb       1  GCA--CTACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
NA_G.sIIb       1  GCA--CCACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
NA_G.cal        1  GCA--CCACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
 C_O.uni        1  GCA--CCACA AG--CGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
CS_O.uni        1  GCA--CCACA AG--CGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
CA_O.uni        1  GCA--CCACA AG--CGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
M_O.uni(V)      1  GCA--CCACA AG--CGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
 C_G.sac        1  GCA--CCACA AG--CGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
CS_G.sac        1  GCA--CCACA AG--CGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
C_G.rub,P       1  GCA--CCACA AG--CGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
NA_G.rub,P      1  GCA--CCACA AG--CGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
NA_G.rub,WI     1  GCA--CCACA AG--CGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
CS_G.rub,WI     1  GCA--CCACA AG--CGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
C_G.rub(P)      1  GCA--CCACA AG--CGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
NA_G.rub,WII    1  GCA--CCACA AG--CGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
C_G.rub,WII     1  GCA--CCACA AG--CGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
C_G.con(P)      1  GCA--CCACA AG--CGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
CS_G.con        1  GCA--CCACA AG--CGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
NA_N.pac(D)     1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
AA_N.pac(D)     1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
 C_N.dut        1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
CS_G.bul,Ia     1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
CA_G.bul,IId    1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
M_G.bul(V)      1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
Na_G.bul,Ib     1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
NA_G.falc       1  GCA--CCACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
CS_G.falc       1  GCA--CCACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
Na_G.bul,IIb    1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
AA_G.bul,IIb    1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
Na-G.bul,IIa    1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
AA_G.bul,IIa    1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
AA_G.bul,IIc    1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
NA-T.qui,IIa    1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TCGACTCAAC
AA_T.qui,IIa    1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TCGACTCAAC
NA_T.qui,IIb    1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TCGACTCAAC
AA_T.qui,IIb    1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TCGACTCAAC
CS_T.qui,I      1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TCGACTCAAC
CS_G.glu        1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
NA_G.uvu        1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
```

```
A.becca_P      1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
 Troch_P       1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
H.germ_P       1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
 Textu_P       1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
 Biger_P       1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
 Boliv_P       1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
G.oper_P       1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
A.trian_P      1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
As.rara_P      1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
A.angu_P       1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
E.acul_P       1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
  Quin_P       1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
M.secan_P      1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
P.pert_P       1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
Allogrom_P     1  GCA--CCACA AGAACGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
Align          1  mmm--mmmmm mm--mmmmm- --mmmmmmm-m mmmmmmmmmmm mmmmmmmmmmm

  C_G.sI      43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
 NA_G.sI      43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
C_G.sI(V)     43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
C_G.sIIa      43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
NA_G.sIIa     43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
CS_G.sIIa     43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
CA_G.sIIa     43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
CA_G.sIIb     43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
NA_G.sIIb     43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
NA_G.cal      43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAT-ACTGA G--GATTG-A
 C_O.uni      43  GCGGG-AAAT CTTACCAGG- -----TCAGG ACAC-ATTGA G--GATTG-A
CS_O.uni      43  GCGGG-AAAT CTTACCAGG- -----TCAGG ACAC-ATTGA G--GATTG-A
CA_O.uni      43  GCGGG-AAAT CTTACTAGA- -----TCAGG ACAC-TTTGA G--GATTG-A
M_O.uni(V)    43  GCGGG-AAAT CTTACTAGA- -----TCAGG ACAC-TTTGA G--GATTG-A
 C_G.sac      43  GCGGG-AAAT CTTACCAGG- -----TCCAG ACAT-ATGGA G--GATTG-A
CS_G.sac      43  GCGGG-AAAT CTTACCAGG- -----TCCAG ACAT-ATGGA G--GATTG-A
C_G.rub,P     43  GCGGG-AAAT CTTACCGGG- -----TCCGG ACAT-ATCGA G--GATTG-A
NA_G.rub,P    43  GCGGG-AAAT CTTACCGGG- -----TCCGG ACAT-ATCGA G--GATTG-A
NA_G.rub,WI   43  ACGGG-AAAT CTTACCGGG- -----TCCGG ACAT-ATCGA G--GATTG-A
CS_G.rub,WI   43  NCGGG-AAAT CTTACCGGG- -----TCCGG ACAT-ATCGA G--GATTG-A
C_G.rub(P)    43  ACGGG-AAAT CTTACCGGG- -----TCCGG ACGT-ATCGA G--GATTG-A
NA_G.rub,WII  43  ACGGG-AAAT CTTACCGGG- -----TCCGG ACAT-ACCGA G--GATTG-A
C_G.rub,WII   43  ACGGG-AAAT CTTACCGGG- -----TCCGG ACAT-ACCGA G--GATTG-A
C_G.con(P)    43  ACGGG-AAAT CTTGCCGGG- -----TCCGG ACAT-ACCGA G--GATTG-A
CS_G.con      43  CGGGG-AAAT CTTACCGGG- -----TCCGG ACAT-ACCGA G--GATTG-A
NA_N.pac(D)   45  GCGGG-AAAT CTTACCAGG- -----TCCGG ACAC-ACTGA G--GATTG-A
AA_N.pac(D)   45  GCGGG-AAAT CTTACCAGG- -----TCCGG ACAC-ACTGA G--GATTG-A
 C_N.dut      45  GCGGG-AAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
CS_G.bul,Ia   43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACAC-AGTGA G--GATTG-A
CA_G.bul,IId  43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACAC-AGTGA G--GATTG-A
M_G.bul(V)    43  GCGGA-AAAG CCTATTTGG- -----TCCGG ACAC-AGTGA G--GATTG-A
Na_G.bul,Ib   43  GCGGA-AAAG CCTATTTGG- -----TCCGG ACAC-AGTGA G--GATTG-A
NA_G.falc     43  GCGGA-AAAA CTTACCAGG- -----TCCGG ACAC-ACTGA G--GATTG-A
CS_G.falc     43  GCGGA-AAAA CTTACCAGG- -----TCCGG ACAC-ACTGA G--GATTG-A
Na_G.bul,IIb  43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACGC-AGTGA G--GATTG-A
AA_G.bul,IIb  43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACGC-AGTGA G--GATTG-A
Na-G.bul,IIa  43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACGC-AGTGA G--GATTG-A
AA_G.bul,IIa  43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACGC-AGTGA G--GATTG-A
AA_G.bul,IIc  43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACAC-AGTGA G--GATTG-A
NA-T.qui,IIa  43  GCGCA-ACAA TTTACTTGG- -----TCCGA ACGC-TTTGG G--GNTTG-A
AA_T.qui,IIa  43  GCGCA-ACAA TTTACTTGG- -----TCCGA ACGC-TTTGG G--GATTG-A
NA_T.qui,IIb  43  GCGCA-ACAA TTTACTTGG- -----TCCGA ACGC-TTTGA G--GATTG-A
AA_T.qui,IIb  43  GCGCA-ACAA TTTACTTGG- -----TCCGA ACGC-TTTGA G--GATTG-A
CS_T.qui,I    43  GCGCA-ATAA CTTACTTGG- -----TCCGA ACGC-TTTGA G--GATTG-A
CS_G.glu      45  GCGGG-AAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
NA_G.uvu      45  GCGGG-AAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
A.becca_P     45  GCGGG-AAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
 Troch_P      45  GCGGG-AAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
H.germ_P      45  GCGGG-AAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
 Textu_P      45  GCGGG-AAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
 Biger_P      45  GCGGG-AAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
 Boliv_P      45  GCGGG-AAAT CTTACCGAG- -----TCCGG ACAC-ACTGA G--GATTG-A
G.oper_P      45  GCGGG-AAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
A.trian_P     45  GCGGG-AAAT CTTACCGGG- -----TCCGG ACAT-ACTGA G--GATTG-A
As.rara_P     45  GCGGG-AAAT CTTACCGGG- -----TCCGG ACAT-ACTGA G--GATTG-A
A.angu_P      45  GCGGG-AAAT CTTACCAGG- -----TCCAG ACAT-ATTGA G--GATTG-A
E.acul_P      45  GCGGG-AAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
  Quin_P      45  GCGGG-AAAT CTTACCAGG- -----TCCAG ACAT-ATTGA G--GATTG-A
M.secan_P     45  GCGGG-AAAT CTTACCAGG- -----TCCGG ACAT-ATTGA G--GATTG-A
P.pert_P      45  GCGGG-AAAT CTTACCAGG- -----TCCGG ACAT-ATTGA G--GATTG-A
```

```
Allogrom_P     45  GCGGG-AAAT CTTACCAGG- -----TCCGA ACAC-GCTGA G--GATTG-A
Align          43  mmmmm-mmmm mmmmmmmmm- -----mmmmm mmmm-mmmmm m--mmmmm-m

  C_G.sI       82  CAGACAGTT- -GTCTTT--- --CCCTCCC- ---------- ----------
 NA_G.sI       82  CAGACAGTT- -GTCTTT--- --CCCTCCC- ---------- ----------
C_G.sI(V)      82  CAGACAGTT- -GTCTTT--- --CCCTCCC- ---------- ----------
C_G.sIIa       82  CAGACAATT- -GTCTTT--T GGTCAAATTT AA-------- ----------
NA_G.sIIa      82  CAGACAATT- -GTCTTT--T TGTCAAATTT AA-------- ----------
CS_G.sIIa      82  CAGACAATT- -GTCTTT--T TGTCAAATTA AA-------- ----------
CA_G.sIIa      82  CAGACAATT- -GTCTTT--T TGTCAAATTT AA-------- ----------
CA_G.sIIb      82  CAGACAATT- -GTCCTT--T TGGCTTAAA- ---------- ----------
NA_G.sIIb      82  CAGACAATT- -GTCCTT--T TGGCTTAAA- ---------- ----------
NA_G.cal       82  CAGACATCA- -CCACGTCTT TTTGTCCTTA AAA------- ----------
 C_O.uni       82  CAGATAGTT- ---------- ---------- --------AT ACCAAAC-AC
CS_O.uni       82  CAGATAGTT- ---------- ---------- --------AT ACCAAAC-AC
CA_O.uni       82  CAGACAGTT- ---------- ---------- --------AT ACCGAACTAC
M_O.uni(V)     82  CAGACAGTT- ---------- ---------- --------AT ACCGAACTAC
 C_G.sac       82  CAGACAGTT- ---------- ---------- --------AA CCAAAACGCA
CS_G.sac       82  CAGACAGTT- ---------- ---------- --------AA CCAAAACGCA
C_G.rub,P      82  CAGACAGTT- --------AC CATG-----C CCTTTCGAA- ----------
NA_G.rub,P     82  CAGACAGTT- --------AC CATG-----C CCTTTCGAA- ----------
NA_G.rub,WI    82  CAGACAGTT- -------ATA CATG-----C CCT--ACAAG T---------
CS_G.rub,WI    82  CAGACAGTT- -------ATA CATG-----C CCT--ACAAG T---------
C_G.rub(P)     82  CAGATAGTT- -------ATA CATG-----C CCT--ACAAG CAAGT-----
NA_G.rub,WII   82  CAGACTGTT- -------AAC TGCAACACAA GCTGCTAA-- ----------
C_G.rub,WII    82  CAGACTGTT- -------AAC TGCAACACAT GCTGCTAA-- ----------
C_G.con(P)     82  CAGACAGTT- ---------- ---------- --------AA CTTGCCCTAC
CS_G.con       82  CAGACAGTT- ---------- ---------- --------AA CTTGCCCTAC
NA_N.pac(D)    84  CAGGAAGTA- ------TCGT CTTTTGAATT CTTTAAGGAC ATGTCGTTTT
AA_N.pac(D)    84  CAGGAAGTA- ------TCGT CTTTTGAATT CTTTAAGGAC ATGTCGTTTT
 C_N.dut       84  CAGGCAATA- ------TCTA AATCGTTTAT AATACTTCCT ATTATAATAC
CS_G.bul,Ia    82  CAGACGGTC- ------TTAT TGGTGGACTC TAAGACGTC- ----------
CA_G.bul,IId   82  CAGACAGTT- ------AGAC AGAAGTGAGG TTTTGGTAAC AA---------
M_G.bul(V)     82  CAGACATGT- ------CGTA TTGGTGGACT TTATGACAAA ----------
Na_G.bul,Ib    82  CAGACATGT- ------CGTA TTGGTGGACT TTATGACAAA ----------
NA_G.falc      82  CAGACGGTT- -----AACGC CCTTCGGGGC AA-------- ----------
CS_G.falc      82  CAGACGGTT- -----GACGC CCTTCGGGGC AAA------- ----------
Na_G.bul,IIb   82  CAGACAGTT- ------AGAC AGAAGTGGTT TAGGTAACAA ----------
AA_G.bul,IIb   82  CAGACAGTT- ------AGAC AGAAGTGGTT TAGGTAACAA ----------
Na-G.bul,IIa   82  CAGACAGTT- ------TCAG GAGTGGTTCT TGGTAAACAA ----------
AA_G.bul,IIa   82  CAGACAGTT- ------TCAG GAGTGGTTCT TGGTAAACAA ----------
AA_G.bul,IIc   82  CAGACAGTT- ------GGCA GGAGTGGTTC TTGGTAAAAA CAA-------
NA-T.qui,IIa   82  CAGTTATTG- ------TATA GTTCTGATAT GAGAGGTCTT TGTAGTCAAC
AA_T.qui,IIa   82  CAGTTATTG- ------TATA GTTCTGATAT GAGAGGTCTT TGTAGTCAAC
NA_T.qui,IIb   82  CAGTTATTG- ------TATA GTTCTGATAT GGGTGGTATT TGTAGTCAAC
AA_T.qui,IIb   82  CAGTTATTG- ------TATA GTTCTGATAT GGGTGGTATT TGTAGTCAAC
CS_T.qui,I     82  CAGTTTCTG- ------GTAC AATGTGCGGC TTGTTTGTTA CAACTACGAA
CS_G.glu       84  CAGGTATTA- ------TATA GCGCGCTTGC GCGTCTA--- ----------
NA_G.uvu       84  CAGGCAATA- ------TATG TGACTCTTCG GAGTTTGCA- ----------
A.becca_P      84  CAGATATAC- ------G-CT CATTGCATGT GCTTCGGCGC TGCTTTG---
 Troch_P       84  CAGGCAATA- ------TTAA TGCTAATCAT AATGCACATT TGTGTATTAT
H.germ_P       84  CAGACAAAT- ------ACAC ACACTTTTAC GGTGTGT--- ----------
 Textu_P       84  CAGGTATTA- ------TTAT ATTGCATTCG ATTTAATCGT ATTGTCAATA
 Biger_P       84  CAGGTATTA- ------TTAT ATAGCGCATA TTTTTAATAT GTTT------
 Boliv_P       84  CAGGTAACA- ------TCTC ACAGTTGCTG CCGTG----- ----------
G.oper_P       84  CAGGTTTTA- ------TCCA TATATTTTAT AT-------- ----------
A.trian_P      84  CAGGTGCAA- ------AAAT GTAATTTATT ATATTCATTT ATAACATATT
As.rara_P      84  CAGGTGCAA- ------AAAT GTATTATAAT TTATAATTTT ATTATTATAT
A.angu_P       84  CAGGCGATA- ------GTTT ATTATATAAT AATA------ ----------
E.acul_P       84  CAGATATAC- ------GTAT ACTATATGTA T--------- ----------
  Quin_P       84  CAGGTGATA- ------ACTA ATATATTTTT AATAT----- ----------
M.secan_P      84  CAGGTGATC- ------ACTA ATATAATTTA TTATATT--- ----------
P.pert_P       84  CAGGCGATA- ------GTAA TATTATTATA TAAT------ ----------
Allogrom_P     84  CAGGTTTTT- ------ATAA GACTATATAT AATTTTTTTT AAAATTATAT
Align          82  mmmmm----- ---------- ---------- ---------- ----------

  C_G.sI      104  ---------T TC---GGGGT GGGA------ ---------- ----------
 NA_G.sI      104  ---------T TC---GGGGT GGGA------ ---------- ----------
C_G.sI(V)     104  ---------T TC---GGGGT GGGA------ ---------- ----------
C_G.sIIa      110  ---------- ---------- ---------- ---------- ----------
NA_G.sIIa     110  ---------- ---------- ---------- ---------- ----------
CS_G.sIIa     110  ---------- ---------- ---------- ---------- ----------
CA_G.sIIa     110  ---------- ---------- ---------- ---------- ----------
CA_G.sIIb     107  ---------- ---------- ---------- ---------- ----------
NA_G.sIIb     107  ---------- ---------- ---------- ---------- ----------
NA_G.cal      113  ---------- ---------- ---------- ---------- ----------
 C_O.uni      102  GAGCTCGAAC CAGAGTTC-- -GTATATGAT ---------- ----------
```

```
CS_O.uni        102  GAGCTCGAAC CAGAGTTC-- -GTATATGAT ---------- ----------
CA_O.uni        103  GAATAATGAA TGGTCCAGGG T--------- ---------- ----------
M_O.uni(V)      103  GAATAATGAA TGGTCCAGGG T--------- ---------- ----------
 C_G.sac        103  GC-------- ---------- AGCTAG-TT- ---------- ----------
CS_G.sac        103  GC-------- ---------- AGCTAG-TT- ---------- ----------
C_G.rub,P       107  ---------- ---------- --AGGAGTCA G--------- ----------
NA_G.rub,P      107  ---------- ---------- --AGGAGTCA G--------- ----------
NA_G.rub,WI     108  ---------- ---------- --AGGAGTCA G--------- ----------
CS_G.rub,WI     108  ---------- ---------- --AGGAGTCA G--------- ----------
C_G.rub(P)      112  ---------- ---------- --AGGAGTCA G--------- ----------
NA_G.rub,WII    112  ---------- ---------- ---------- ---------- ----------
C_G.rub,WII     112  ---------- ---------- ---------- ---------- ----------
C_G.con(P)      103  CTATAGGAGC TAG------- ---------- ---------- ----------
CS_G.con        103  CTATAGGAGC TAG------- ---------- ---------- ----------
NA_N.pac(D)     127  TAATGACATC TTTTAGATGG ATGATTC--- ---------- ----------
AA_N.pac(D)     127  TAATGACATC TTTTAGATGG ATGATTC--- ---------- ----------
 C_N.dut        127  GCATTTA--- ---------- ---------- ---------- ----------
CS_G.bul,Ia     114  ---------- ---------- ---------- ---------- ----------
CA_G.bul,IId    117  ---------- ---------- ---------- ---------- ----------
M_G.bul(V)      115  ---------- ---------- ---------- ---------- ----------
Na_G.bul,Ib     115  ---------- ---------- ---------- ---------- ----------
NA_G.falc       108  ---------- ---------- ---------- ---------- ----------
CS_G.falc       109  ---------- ---------- ---------- ---------- ----------
Na_G.bul,IIb    115  ---------- ---------- ---------- ---------- ----------
AA_G.bul,IIb    115  ---------- ---------- ---------- ---------- ----------
Na-G.bul,IIa    115  ---------- ---------- ---------- ---------- ----------
AA_G.bul,IIa    115  ---------- ---------- ---------- ---------- ----------
AA_G.bul,IIc    118  ---------- ---------- ---------- ---------- ----------
NA-T.qui,IIa    125  GTGTAGGTAG TTGTA----- ---------- ---------- ----------
AA_T.qui,IIa    125  GTGTAGGTAG TTGTA----- ---------- ---------- ----------
NA_T.qui,IIb    125  GTGTAGGTAG TTGTA----- ---------- ---------- ----------
AA_T.qui,IIb    125  GTGTAGGTAG TTGTA----- ---------- ---------- ----------
CS_T.qui,I      125  TGATTCGAAT TGTTGTAAAT ATGACTTGGC CGGCCTTCGG GTGTCTTGGA
CS_G.glu        114  ---------- ---------- ---------- ---------- ----------
NA_G.uvu        116  ---------- ---------- ---------- ---------- ----------
A.becca_P       123  ---------- ---------- ---------- ---------- ----------
 Troch_P        127  TGATTTAGTA TT-------- ---------- ---------- ----------
H.germ_P        114  ---------- ---------- ---------- ---------- ----------
 Textu_P        127  T--------- ---------- ---------- ---------- ----------
 Biger_P        121  ---------- ---------- ---------- ---------- ----------
 Boliv_P        112  ---------- ---------- ---------- ---------- ----------
G.oper_P        109  ---------- ---------- ---------- ---------- ----------
A.trian_P       127  ---------- ---------- ---------- ---------- ----------
As.rara_P       127  TTTATTAAT- ---------- ---------- ---------- ----------
A.angu_P        111  ---------- ---------- ---------- ---------- ----------
E.acul_P        108  ---------- ---------- ---------- ---------- ----------
  Quin_P        112  ---------- ---------- ---------- ---------- ----------
M.secan_P       114  ---------- ---------- ---------- ---------- ----------
P.pert_P        111  ---------- ---------- ---------- ---------- ----------
Allogrom_P      127  ATAGCAATTC TT-------- ---------- ---------- ----------
 Align           87  ---------- ---------- ---------- ---------- ----------

      C_G.sI    116  ---------- ---------- ---------- --TTACAAAA GATAACA-GA
     NA_G.sI    116  ---------- ---------- ---------- --TTACAAAA GATAACA-GA
   C_G.sI(V)    116  ---------- ---------- ---------- --TTACAAAA GATAACA-GA
    C_G.sIIa    110  ---------- ---------- ---------- --AGACAAAA GATAATA-GA
   NA_G.sIIa    110  ---------- ---------- ---------- --AGACAAAA GATAACA-GA
   CS_G.sIIa    110  ---------- ---------- ---------- --AGACAAAA GATAACA-GA
   CA_G.sIIa    110  ---------- ---------- ---------- --AGACAAAA GATAACA-GA
   CA_G.sIIb    107  ---------- ---------- ---------- --AGACAAAA GATAACA-GA
   NA_G.sIIb    107  ---------- ---------- ---------- --AGACAAAA GATAACA-GA
    NA_G.cal    113  ---------- ---------- ---------- --AGACAAAA GATAACA-GA
     C_O.uni    129  ---------- ---------- ---------- --TTACAAAA TCTAACG-GC
    CS_O.uni    129  ---------- ---------- ---------- --TTACAAAA TCTAACG-GC
    CA_O.uni    124  ---------- ---------- ---------- --AGGCAAAA ACTAATA-GC
  M_O.uni(V)    124  ---------- ---------- ---------- --AGGCAAAA ACTAATA-GC
     C_G.sac    113  ---------- ---------- ---------- ---TAACAAA ACTAACA-GC
    CS_G.sac    113  ---------- ---------- ---------- ---TAACAAA ACTAACA-GC
   C_G.rub,P    116  ---------- ---------- ---------- --GTTTTAAA CCTAATG-GC
  NA_G.rub,P    116  ---------- ---------- ---------- --GTTTTAAA CCTAATG-GC
 NA_G.rub,WI    117  ---------- ---------- ---------- --GTTTTAAA CCTAATG-GC
 CS_G.rub,WI    117  ---------- ---------- ---------- --GTTTTAAA CCTAATG-GC
  C_G.rub(P)    121  ---------- ---------- ---------- --GTTTTAAA CCTAATG-GC
NA_G.rub,WII    112  ---------- ---------- ---------- --GTTTTAAA CTTAATA-GC
 C_G.rub,WII    112  ---------- ---------- ---------- --GTTTTAAA CTTAATA-GC
  C_G.con(P)    116  ---------- ---------- ---------- --GTTCTAAA CCTAATA-GC
    CS_G.con    116  ---------- ---------- ---------- --GTTCTAAA CCTAATA-GC
```

```
NA_N.pac(D)    154  ---------- ---------- ---------- --GTG-TAAA TATGCTA-GT
AA_N.pac(D)    154  ---------- ---------- ---------- --GTG-TAAA TATGCTA-GT
    C_N.dut    134  ---------- ---------- ---------- --GTGTTAAA TATGCTA-GT
CS_G.bul,Ia    114  ---------- ---------- ---------- ---ATGGTAA AGTTGAA-GT
CA_G.bul,IId   117  ---------- ---------- ---------- ---TTGAGAG AGTTGAA-GT
  M_G.bul(V)   115  ---------- ---------- ---------- ---ATGGTAA AGTTGAA-GT
 Na_G.bul,Ib   115  ---------- ---------- ---------- ---ATGGTAA AGTTGAA-GT
   NA_G.falc   108  ---------- ---------- ---------- --GAGACAAA ATTTTTG-GC
   CS_G.falc   109  ---------- ---------- ---------- --GAGACAAA ATTTTTG-GC
Na_G.bul,IIb   115  ---------- ---------- ---------- --TTG-AGAG AGTTGAA-GT
 AA_G.bul,IIb  115  ---------- ---------- ---------- --TTG-AGAG AGTTGAA-GT
Na-G.bul,IIa   115  ---------- ---------- ---------- --TTG-AGAG AGTTGAA-GT
 AA_G.bul,IIa  115  ---------- ---------- ---------- --TTG-AGAG AGTTGAA-GT
 AA_G.bul,IIb  118  ---------- ---------- ---------- --ATG-AGAG AGTTGAA-GT
NA-T.qui,IIa   140  ---------- ---------- ---------- --TAT-TAAA TATGAAA-GT
 AA_T.qui,IIa  140  ---------- ---------- ---------- --TAT-TAAA TATGAAA-GT
NA_T.qui,IIb   140  ---------- ---------- ---------- --TAT-TAAA TATGAAA-GT
 AA_T.qui,IIb  140  ---------- ---------- ---------- --TAT-TAAA TATGAAA-GT
  CS_T.qui,I   175  TCGGCGTATC AGGTCGTACA TTG------- --TAT-TCAA TGAGAAA-GT
    CS_G.glu   114  ---------- ---------- ---------- --TTGTTAAA TATGCTA-GT
    NA_G.uvu   116  ---------- ---------- ---------- --TTGTTAAA TATGCTA-GT
   A.becca_P   123  ---------- ---------- ---------- --AGCTGAAA GATGCTA-GT
     Troch_P   139  ---------- ---------- ---------- --ATGTTGAA TATGCTA-GT
     H.germ_P  114  ---------- ---------- ---------- --GCATGAAA TATGCTA-GT
     Textu_P   128  ---------- ---------- ---------- --ATGTTAAA TATGCTA-GT
     Biger_P   121  ---------- ---------- ----GTCTAT ATATGTTAAA TATGCTA-GT
     Boliv_P   112  ---------- ---------- ---------- --GGTATAAA TATGCTA-GT
    G.oper_P   109  ---------- ---------- ---------A TGGTGTCAAA AATGCTA-GT
   A.trian_P   127  ---------- ---------- ---------A CATTATCAAA TATGCTA-GT
   As.rara_P   136  ---------- ---------- ---------A CGTTATCAAA TATGCTA-GT
    A.angu_P   111  ---------- ---------- ---------- --AGCATAAA AATGATA-GT
    E.acul_P   108  ---------- ---------- ---------- -ACGTTCAAA GATGCTA-GT
      Quin_P   112  ---------- ---------- ---------A TTAGTACAAA TATGATA-GT
   M.secan_P   114  ---------- ---------- ---------- --AGTACAAA TATGATA-GT
    P.pert_P   111  ---------- ---------- ---------- ATTGCATAAA AATGATA-GT
  Allogrom_P   139  ---------- ---------- ---------- --ATATCAAA TATGCTA-GT
       Align    87  ---------- ---------- ---------- ---------- --mmmmm-mm

      C_G.sI    133  ---------- ---------T CTTTCATGA- -----TCATG TGATGG----
     NA_G.sI    133  ---------- ---------T CTTTCATGA- -----TCATG TGATGG----
   C_G.sI(V)    133  ---------- ---------T CTTTCATGA- -----TCATG TGATGG----
    C_G.sIIa    127  ---------- ---------T CTTTCATGA- -----TTATG TGATGG----
   NA_G.sIIa    127  ---------- ---------T CTTTCATGA- -----TTATG TGATGG----
   CS_G.sIIa    127  ---------- ---------T CTTTCATGA- -----TTATG TGATGG----
   CA_G.sIIa    127  ---------- ---------T CTTTCATGA- -----TTATG TGATGG----
   CA_G.sIIb    124  ---------- ---------T CTTTCATGA- -----TTATG TGATGG----
   NA_G.sIIb    124  ---------- ---------T CTTTCATGA- -----TTATG TGATGG----
    NA_G.cal    130  ---------- ---------T CTTTCATGA- -----TTATG TGGTGG----
    C_O.uni     146  ---------- ---------T CTTACATGA- -----TTATG TGATAG----
   CS_O.uni     146  ---------- ---------T CTTACATGA- -----TTATG TGATAG----
   CA_O.uni     141  ---------- ---------T CTTTTATGA- -----TTATG TGATAG----
  M_O.uni(V)    141  ---------- ---------T CTTTTATGA- -----TTATG TGATAG----
    C_G.sac     129  ---------- ---------T CTTTCAAGA- -----TTATG TGGTAG----
   CS_G.sac     129  ---------- ---------T CTTTCAAGA- -----TTATG TGGTAG----
   C_G.rub,P    133  ---------- ---------T CTTTCATGA- -----TTATA TGATAG----
  NA_G.rub,P    133  ---------- ---------T CTTTCATGA- -----TTATA TGATAG----
  NA_G.rub,WI   134  ---------- ---------T CTTTCATGA- -----TTATA TGATAG----
  CS_G.rub,WI   134  ---------- ---------T CTTTCATGA- -----TTATA TGATAG----
   C_G.rub(P)   138  ---------- ---------T CTTTCATGA- -----TTATA TGATAG----
 NA_G.rub,WII   129  ---------- ---------T CTTTCATGA- -----TTATA TGATAG----
  C_G.rub,WII   129  ---------- ---------T CTTTCATGA- -----TTATA TGATAG----
   C_G.con(P)   133  ---------- ---------T CTTTCATGA- -----TTATA TGATAG----
    CS_G.con    133  ---------- ---------T CTTTCATGA- -----TTATA TGATAG----
  NA_N.pac(D)   170  ---------- ---------T CTTTCATGA- -----TTATG TGATAG----
  AA_N.pac(D)   170  ---------- ---------T CTTTCATGA- -----TTATG TGATAG----
    C_N.dut     151  ---------- ---------C CTTTCATGA- -----TTATG TGATAG----
 CS_G.bul,Ia    130  ---------- ---------T CTTTCATGA- -----TCTTG TGGCAG----
CA_G.bul,IId    133  ---------- ---------T CTTTCATGA- -----TCTTG TGAGAG----
  M_G.bul(V)    131  ---------- ---------T CTTTCATGA- -----TCTTG TGGTAG----
 Na_G.bul,Ib    131  ---------- ---------T CTTTCATGA- -----TCTTG TGGTAG----
   NA_G.falc    125  ---------- ---------T CTTTCATGA- -----TTATG TGATAG----
   CS_G.falc    126  ---------- ---------T CTTTCATGA- -----TTATG TGATAG----
Na_G.bul,IIb    131  ---------- ---------T CTTTCATGA- -----TCTTG TGAGAG----
 AA_G.bul,IIb   131  ---------- ---------T CTTTCATGA- -----TCTTG TGAGAG----
Na-G.bul,IIa    131  ---------- ---------T CTTTCATGA- -----TCTTG TGAGAG----
 AA_G.bul,IIa   131  ---------- ---------T CTTTCATGA- -----TCTTG TGAGAG----
 AA_G.bul,IIc   134  ---------- ---------T CTTTCATGA- -----TCTTG TGAGAG----
```

```
NA-T.qui,IIa   156  ---------- ---------T CTTTTATGA- -----TTATG TGATAG----
  AA_T.qui,IIa 156  ---------- ---------T CTTTTATGA- -----TTATG TGATAG----
NA_T.qui,IIb   156  ---------- ---------T CTTTTATGA- -----TTATG TGATAG----
  AA_T.qui,IIb 156  ---------- ---------T CTTTTATGA- -----TTATG TGATAG----
  CS_T.qui,I   214  ---------- ---------T CTTTTATGA- -----TTATG TGGTAG----
    CS_G.glu   131  ---------- ---------C CTTTCATGA- -----TTATG TGATAG----
    NA_G.uvu   133  ---------- ---------C CTTTCATGA- -----TTATG TGATAG----
   A.becca_P   140  ---------- ---------T CTTTCATGA- -----TTATG TGATAG----
     Troch_P   156  ---------- ---------C CTTTCATGA- -----TTATG TGATAG----
     H.germ_P  131  ---------- ---------T CTTTCATGA- -----TTATG TGATAG----
      Textu_P  145  ---------- ---------C CTTTCATGA- -----TTATG TGATAG----
      Biger_P  146  ---------- ---------C CTTTCATGA- -----TTATG TGATAG----
      Boliv_P  129  ---------- ---------C CTTTCATGA- -----TTATG TGATAG----
     G.oper_P  129  ---------- ---------C CTTTCATGA- -----TTATG TGATAG----
    A.trian_P  147  ---------- ---------C CTTTCATGA- -----TTGTG TGATAG----
    As.rara_P  156  ---------- ---------C CTTTCATGA- -----TTGTG TGATAG----
     A.angu_P  128  ---------- ---------C CTTTCATGA- -----TTATA TGATAG----
     E.acul_P  126  ---------- ---------T CTTTCATGA- -----TTATG TGATAG----
       Quin_P  132  ---------- ---------C CTTTCATGA- -----TTATA TGATAG----
    M.secan_P  131  ---------- ---------C CTTTCATGA- -----TTATA TGATAG----
     P.pert_P  130  ---------- ---------C CTTTCATGA- -----TTATA TGATAG----
   Allogrom_P  156  ---------- ---------C CTTTCATGA- -----TTGCG TGATAG----
        Align   94  ---------- ---------m mmmmmmmmm- -----mmmmm mmmmmm----

       C_G.sI   154  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
      NA_G.sI   154  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
    C_G.sI(V)   154  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
     C_G.sIIa   148  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
    NA_G.sIIa   148  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
    CS_G.sIIa   148  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
    CA_G.sIIa   148  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
    CA_G.sIIb   145  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
    NA_G.sIIb   145  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
     NA_G.cal   151  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
      C_O.uni   167  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
     CS_O.uni   167  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
     CA_O.uni   162  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
    M_O.uni(V)  162  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
      C_G.sac   150  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GAA-GTGATT -CGTT--CGC
     CS_G.sac   150  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GAA-GTGATT -CGTT--CGC
    C_G.rub,P   154  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
   NA_G.rub,P   154  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
  NA_G.rub,WI   155  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
  CS_G.rub,WI   155  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
    C_G.rub(P)  159  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
 NA_G.rub,WII   150  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
  C_G.rub,WII   150  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
    C_G.con(P)  154  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
      CS_G.con  154  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
   NA_N.pac(D)  191  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
   AA_N.pac(D)  191  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
      C_N.dut   172  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
   CS_G.bul,Ia  151  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GTA-GTGATA -CGTC--TGC
  CA_G.bul,IId  154  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GTA-GTGATA -TGTC--TGC
    M_G.bul(V)  152  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GTA-GTGATA -CGTC--TGC
   Na_G.bul,Ib  152  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GTA-GTGATA -CGTC--TGC
     NA_G.falc  146  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
     CS_G.falc  147  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
  Na_G.bul,IIb  152  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GTA-GTGATA -TGTC--TGC
   AA_G.bul,IIb 152  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GTA-GTGATA -TGTC--TGC
  Na-G.bul,IIa  152  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GTA-GTGATA -TGTC--TGC
   AA_G.bul,IIa 152  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GTA-GTGATA -TGTC--TGC
   AA_G.bul,IIc 155  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GTA-GTGATA -TGTC--TGC
NA-T.qui,IIa   177  -GTGGTG-CA TGG-CCGTC- CTTAATTCGT GGA-GTGATT -TGTC--TGC
  AA_T.qui,IIa 177  -GTGGTG-CA TGG-CCGTC- CTTAATTCGT GGA-GTGATT -TGTC--TGC
NA_T.qui,IIb   177  -GTGGTG-CA TGG-CCGTC- CTTAATTCGT GGA-GTGATT -TGTC--TGC
  AA_T.qui,IIb 177  -GTGGTG-CA TGG-CCGTC- CTTAATTCGT GGA-GTGATT -TGTC--TGC
  CS_T.qui,I   235  -GTGGTG-CA TGG-CCGTC- TTTAATTCGT GGA-GTGATC -TGTC--TGC
    CS_G.glu   152  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
    NA_G.uvu   154  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
   A.becca_P   161  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
     Troch_P   177  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
     H.germ_P  152  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
      Textu_P  166  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
      Biger_P  167  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
      Boliv_P  150  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
     G.oper_P  150  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
```

```
A.trian_P   168  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
As.rara_P   177  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
 A.angu_P   149  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATT -TGTC--TGC
 E.acul_P   147  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
    Quin_P  153  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
M.secan_P   152  -GTGGTG-CG TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
 P.pert_P   151  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATT -TGTC--TGC
Allogrom_P  177  -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT GGA-GTGATC -TGTC--TGC
     Align  115  -mmmmmm-mm mmm-mmmmm- mmmmmmmmmmm mmm-mmmmmm -mmmm--mmm

      C_G.sI  196  TT-AATTGCG TTTC------ -AAATTTTAT ATTTGTTAAC AT-GCGAGTC
     NA_G.sI  196  TT-AATTGCG TTTC------ -AAATTTTAT ATTTGTTAAC AT-GCGAGTC
   C_G.sI(V)  196  TT-AATTGCG TTTC------ -AAATTTTAT ATTTGTTAAC AT-GCGAGTC
     C_G.sIIa 190  TT-AATTGCG TTTC------ -AAATATATT TTATTATATA AC-GAGTCTA
    NA_G.sIIa 190  TT-AATTGCG TTTC------ -AAATATATT TTATTATATA AC-GAGTCTA
    CS_G.sIIa 190  TT-AATTGCG TTTC------ -AAATATATT TTATTATATA -C-GAGTCTA
    CA_G.sIIa 190  TT-AATTGCG TTTC------ -AAATATATT TTATTATATA AC-GAGTCTA
    CA_G.sIIb 187  TT-AATTGCG TTTC------ -AAATATACT TTTTTTACAA -C-GAGTCTA
    NA_G.sIIb 187  TT-AATTGCG TTTC------ -AAATATACT TTTTTTACAA -C-GAGTCTA
     NA_G.cal 193  TT-AATTGCG TTTC------ -AAATGTATT TTTGATTTCT ----------
       C_O.uni 209  TT-AATTGCG TTTG------ -GATACAACA ATTTT----- ---GGAATCG
      CS_O.uni 209  TT-AATTGCG TTTG------ -GATACAACA ATTTT----- ---GGAATCG
      CA_O.uni 204  TT-AATTGCG TTTC------ -AAAACAACT GATACACAGA GTCTTGGCAC
    M_O.uni(V) 204  TT-AATTGCG TTTC------ -AAAACAACT GATACACAGA GTCTTGGCAC
       C_G.sac 192  CC-AATCGCA TTTC------ CAAA-CAACT AACAACTCA- ---GAATGCT
      CS_G.sac 192  CC-AATCGCA TTTC------ CAAA-CAACT AACAACTCA- ---GAATGCT
     C_G.rub,P 196  TT-AATTGCG TTTC------ -AAATTTT-T ATTT------ ---GAGGACC
    NA_G.rub,P 196  TT-AATTGCG TTTC------ -AAATTTT-T ATTT------ ---GAGGACC
   NA_G.rub,WI 197  TT-AATTGCG TTTC------ -AAAATTT-A ACTT------ ---GAGGACC
   CS_G.rub,WI 197  TT-AATTGCG TTTC------ -AAAATTT-A ACTT------ ---GAGGACC
     C_G.rub(P) 201  TT-AATTGCG TTTC------ -AAAATTT-A ACTT------ ---GAGGACC
  NA_G.rub,WII 192  TT-AATTGCG TTTC------ -AAAATTTAA ACCT------ ---GAGCGAC
   C_G.rub,WII 192  TT-AATTGCG TTTC------ -AAAATCTAA ACCT------ ---GAGCGAC
     C_G.con(P) 196  TT-AATTGCG TTTC------ -AAAATATGA ATTTGAGTAC CAGGTCCCAA
       CS_G.con 196  TT-AATTGCG TTTC------ --AAATATGA ATTTGAGTAC CAGGTCCCAA
   NA_N.pac(D) 233  TT-AATTGCG TTTC------ -ACTAAGGCC CCAAAGTTAG CAAATTATCA
   AA_N.pac(D) 233  TT-AATTGCG TTTC------ -ACTAAGGCC CCAAAGTTAG CAAATTATCA
       C_N.dut 214  TT-AATTGCG TTTC------ -ACTAAGGGC CCATAAATTC AAGGTATGTT
    CS_G.bul,Ia 193  CT-AATCGCG TCAC------ -GATAACCTA TTTCTCGACA ATCCAAAGAT
   CA_G.bul,IId 196  CT-AATCGCG TCAC------ -GATAACCTA TTGGTCGACA ACCCAATTAT
     M_G.bul(V) 194  CT-AATCGCG TCAC------ -GATAAGCTA TTTCTCGACA ATCCAAAAAT
    Na_G.bul,Ib 194  CT-AATCGCG TCAC------ -GATAAGCTA TTTCTCGACA ATCCAAAAAT
      NA_G.falc 188  CC-AATCGCG TTTC------ -ATAAATCTT AAGCTCAAAA GGCAGTCTTA
      CS_G.falc 189  CC-AATCGCG TTTC------ -ATAAATCTT AAGCTCAAAA GGCAGTCTTA
   Na_G.bul,IIb 194  CT-AATCGCG TCAC------ -GATAACCTA TTGGTCGACA ACCCAATTAT
   AA_G.bul,IIb 194  CT-AATCGCG TCAC------ -GATAACCTA TTGGTCGACA ACCCAATTAT
   Na-G.bul,IIa 194  CT-AATCGCG TCAC------ -GATAACCTA TTGGTCGACA ACCCAATTAT
   AA_G.bul,IIa 194  CT-AATCGCG TCAC------ -GATAACCTA TTGGTCGACA ACCCAATTAT
   AA_G.bul,IIc 197  CT-AATCGCG TCAC------ -GATAACCTA TTGGTCGACA ACCCAATTAT
   NA-T.qui,IIa 219  TT-AATTGCG CATT------ -GCAAATTGT AATTGATCTT TTACCAGTCT
   AA_T.qui,IIa 219  TT-AATTGCG CATT------ -GCAAATTGT AATTGATCTT TTACCAGTCT
   NA_T.qui,IIb 219  TT-AATTGCG CATT------ -GCAAATTGT AATTGATCTT TTACCAGTCT
   AA_T.qui,IIb 219  TT-AATTGCG CATT------ -GCAAATTGT AATTGATCTT TTACCAGTCT
     CS_T.qui,I 277  TT-AATTGCG CATT------ -GCAAATTGT AATTGATCTT TGAACAGATC
       CS_G.glu 194  TT-AATTGCG TTTC------ -AGATAAGGA TCTATATTCT TTACAGAAGT
       NA_G.uvu 196  TT-AATTGCG TTTC------ -AAATTACTT ATAAGGTAGC AATACGCGAT
      A.becca_P 203  TT-AATTGCG TATC------ -AATAATAGA GACCTAGTAT ACGCGTGGAT
        Troch_P 219  TT-AATTGCG TTTC------ -ACTAAGGGC TTATAAATTA CGTGTGTTGC
       H.germ_P 194  TT-AATTGCG TTCC------ -ACTAAAGGC ACATACACAT AATATTGATG
        Textu_P 208  TT-AATTGCG TTTC------ -ACTAAGGGC CTATAAATTA CGTGTGTTGC
        Biger_P 209  TT-AATTGCG TTTC------ -ACTAAGGGC CTATATATCA CGTGTGTTGC
        Boliv_P 192  TT-AATTGCG TTTC------ -ACTATTGGA TCTATATACC GTGCATGCGT
       G.oper_P 192  TT-AATTGCG TTTC------ -ACTACGAAT CTACTTTAAA CGTGTGTTTG
      A.trian_P 210  TT-AATTGCG TTTC------ -ATAATAGCT CCTTATAGAC TTTTCTATTG
      As.rara_P 219  TT-AATTGCG TTTC------ -ATAATAGCT CCAAATAGAC TTTTCTATTG
       A.angu_P 191  TT-AATTGCG TTTC------ -AGTAAAATA AAATATATTT AATATATAAT
       E.acul_P 189  TT-AATTGCG TTTC------ -ATATATACG TACATATTTT CATATGT---
        Quin_P  195  TT-AATTGCG TTTC------ -ATATTATAT TAAATAATAT TATTCAATAT
      M.secan_P 194  TT-AATTGCG TTTC------ -AGTAAAAAA TATATAATTT AAATTTTTAT
       P.pert_P 193  TT-AATTGCG TTTC------ -AGATATATA ATTATATAAT ATATTATATA
     Allogrom_P 219  TT-AACTGCG TTTC------ -ACTATAATG AGTATATATT GAATACTTTG
          Align 157  mm-mmmmmmm mmmm------ ---------- ---------- ----------

      C_G.sI  237  TTCAATGACG GCCGCTCCAC CTTATAGACA TTATATGATA TGCACATTTG
     NA_G.sI  237  TTCAATGACG GCCGCTCCAC CTTATAGACA TTATATGATA TGCACATTTG
   C_G.sI(V)  237  TTCAATGACG GCCGCTCCAC CTTATAGACA TTATATGATA TGCACATTTG
     C_G.sIIa 231  CAAA-----G GAGTGGTTCA TTAACAGACA ATCATATCTT TGAGCGTCTG
```

```
   NA_G.sIIa   231  CAAA-----G GAGTGGTTCA TTAACAGACA ATCATATCTT TGAGCGTCTG
   CS_G.sIIa   230  CAAA-----G GAGTGGTTCA TTAACAGACA ATCATATCTT TGAGCGTCTG
   CA_G.sIIa   231  CAAA-----G GAGTGGTTCA TTAACAGACA ATCATATCTT TGAGCGTCTG
   CA_G.sIIb   227  CAAA-----G GAGTGGTTCA GTTACAACAG ACAATTAATT CTTGTGTTTT
   NA_G.sIIb   227  CAAA-----G GAGTGGTTCA GTTACAACAG ACAATTAATT CTTGTGTTTT
    NA_G.cal   225  CAAA-----C GAGTCTACGA AGATGTCTAT TCATAACAAG CACTTCATTT
      C_O.uni  243  CGGCAATGTT CAGCTTAATG GCGTCGATAA AATTAACATT ATGTTGACAA
     CS_O.uni  243  CGGCAATGTT CAGCTTAATG GCGTCGATAA AATTAACATT ATGTTGACAA
     CA_O.uni  246  AGTTCAGTTT AATGGCGTCG AGTATCTTTG ATTTCGACAG TAGTCCAACT
   M_O.uni(V)  246  AGTTCAGTTT AATGGCGTCG AGTATCTTTG ATTTCGACAG TAGTCCAACT
      C_G.sac  230  GGGCAATAGG ACGTTTAAAG TGGCTCCTTA CACCTTACAA AGGAATGGCT
     CS_G.sac  230  GGGCAATAGG ACGTTTAAAG TGGCTCCTTA CACCTTACAA AGGAATGGCT
   C_G.rub,P   228  AGGTTTTGGT TTGGGTAGTG GTCCTACGAT GACTGTGAAC TGTCATGTCT
  NA_G.rub,P   228  AGGTTTTGGT TTGGGTAGTG GTCCTACGAT GACTGTGAAC TGTCATGTCT
 NA_G.rub,WI   229  GGGTTTTGGT TTGGCTAGTG GTTCTACGAT GACTGTGAAC TGTCATGTCT
 CS_G.rub,WI   229  GGGTTTTGGT TTGGCTAGTG GTTCTACGAT GACTGTGAAC TGTCATGTCT
   C_G.rub(P)  233  GGGTTTTGGT TTGGCTAGTG GTCCTACGAT GACTGTGAAC TGTCATGTCT
NA_G.rub,WII   225  AAGGTGCTGC TATCGAAACT GTCTTTGAGT GTGAACTATA AGTCTTTTGA
 C_G.rub,WII   225  AAGGTGCTGC TATCGAAACT GTCTTTGAGT GTGAACTATA AGTCTTTTGA
   C_G.con(P)  238  GTGTTTCAAG TAAATCTGTC TCTGAGTGTG AACTGTATCT CTTCGGATAT
     CS_G.con  237  GTGTTTCAAG TAAATCTGTC TCTGAGTGTG AACTGTATCT CTTCGGATAT
   NA_N.pac(D) 275  ATCGTTACAG AGTCGACCCC TCACCTTTGA GTGCGCGTCC TAACTTGTAG
   AA_N.pac(D) 275  ATCGTTACAG AGTCGACCCC TCACCTTTGA GTGCGCGTCC TAACTTGTAG
      C_N.dut  256  AGCTATCGTT TCTCAATTGA CCCCTTGTCT TCGATAAGCG CGTGTCTTTT
 CS_G.bul,Ia   235  CTTCGTGCAG TA-TCTATTG GTACGGTGCT CACGTTAGTA GGCTAGAG--
CA_G.bul,IId   238  CACAACTGCA GCATAACTCC CCTTGGGTGG GCGAGGCTCT GTGTAGGATA
   M_G.bul(V)  236  CACGGCGGGA GGTTCTACTG GTACGGGTTC CCAAAGTGTA GGCTAGAG--
  Na_G.bul,Ib  236  CACGGCGGGA GGTTCTACTG GTACGGGTTC CCAAAGTGTA GGCTAGAG--
     NA_G.falc 230  CCTTCGGGTT GAGTTTTGCT GGGAGA---- ---------- ----------
     CS_G.falc 231  CCTTCGGGTT GAGTTTTGCT GGGAGA---- ---------- ----------
  Na_G.bul,IIb 236  CACAACTGCA GCATAACTCC CCTTGGGTGG GCGAGGCTCT GTGTAGGATA
  AA_G.bul,IIb 236  CACAACTGCA GCATAACTCC CCTTGGGTGG GCGAGGCTCT GTGTAGGATA
 Na-G.bul,IIa  236  CACAACTGCA GCATAACTCC CCTTGGGTGG GCGAGGCTCT GTGTAGGATA
  AA_G.bul,IIa 236  CACAACTGCA GCATAACTCC CCTTGGGTGG GCGAGGCTCT GTGTAGGATA
  AA_G.bul,IIc 239  CACAACTGCA GCATAACTCC CCTTGGGTGG GCGAGGCTCT GTGTAGGATA
 NA-T.qui,IIa  261  CGTTCCGATT GATGTTGTGA ATATTGTAGA ATATTGTTGT CGTCGAGTCA
  AA_T.qui,IIa 261  CGTTCCGATT GATGTTGTGA ATATTGTAGA ATATTGTTGT CGTCGAGTCA
  NA_T.qui,IIb 261  CGTTCTGATT TATGTTGTAA TATTGTAGAA TATTGTTGTC GTCGAGTCAT
  AA_T.qui,IIb 261  CGTTCTGATT TATGTTGTAA TATTGTAGAA TATTGTTGTC GTCGAGTCAT
   CS_T.qui,I  319  CGTCTTTTTA TGTTGAATAG TACGTACAGT CAGACCCGGT TGGGCATGTA
     CS_G.glu  236  GTTGCGTAGT GCATATGACC CCTCGTTCGC GAGTGCGTGT CTTTTCGCGC
     NA_G.uvu  238  GAAGCTTTCA ATAGTCCTCA CGGATTCACT TGAGTCTCTC GCTTCGC---
    A.becca_P  245  TT------CG CGTGGTAGTG ACCCCCTGTT TCAACGCAGG CGTGTGTCGC
      Troch_P  261  ATGTACTTTG ACCCCT-AAT CTGAAATATG --GTTGGTGC GTGTCTTAGT
     H.germ_P  236  CGTTGTGTGT ATAATTTATA ATTATACGCA TACACACATA TTATA-----
      Textu_P  250  AGGTACTTTG ACCCCTTTGG TTGAAATATT ACTAAAGTGC GTGTCTTAGT
      Biger_P  251  AGGTACTTTG ACCCCTTTGG TTGAAATATT ACTAAAGTGC GTGTCTTAGT
      Boliv_P  234  TGCGGCATTG ACCCCATTGA TTCACTTCAA TGCGCGCGTC TTTCGCTTAG
     G.oper_P  234  TGAGATGGAC TGATCCCTCC CACCTCTCTG AGGTCTGGCG AGTTGTGCGT
    A.trian_P  252  ATATCAGCCT TAATACTTGA GAATGATCGT GTTATATATA TATATATATA
    As.rara_P  261  ATATCAGCCT TATCACTTTT GAATTTATAA TATATATTTA ATCATAATTT
     A.angu_P  233  AATATTTTAG TTCTGCCTTT ATGGATTTAA AGTGAACATA TTATTATTAT
     E.acul_P  228  ---------- ---------- ---------- ---------- ----------
       Quin_P  237  AATATTATTA GCGTTCTGCC TTTTATAGGA TTGTGAACCT TTTATATTAT
    M.secan_P  236  ATATTTTGAC GTTCTGCCTT ATTTTTAAGG ATTGTGAACT CAAATATATT
     P.pert_P  235  GTAATATATA ATTTAATATT GTGCTGCCTT ATATATTATT TATAAGGATT
   Allogrom_P  261  TTTGCACATA AAGTTGCTGC ATTGTTTTTT AACTTTGCAC CTTTATTGTT
        Align  170  ---------- ---------- ---------- ---------- ----------

      C_G.sI   287  TGTATTTGAT TATAACTTGT CTGGA--GTC TGG-CTCGAT TTTTTT----
      NA_G.sI  287  TGTATTTGAT TATAACTTGT CTGGA--GTC TGG-CTCGAT TTTTTT----
    C_G.sI(V)  287  TGTATTTGAT TATAACTTGT CTGGA--GTC TGG-CTCGAT TTTTTT----
     C_G.sIIa  276  G-AATCTACT CTATTT---- ---------- ---------- ----------
    NA_G.sIIa  276  G-AATCTACT CTATTT---- ---------- ---------- ----------
    CS_G.sIIa  275  G-AATCTACT CTATTT---- ---------- ---------- ----------
    CA_G.sIIa  276  G-AATCTACT CTATTT---- ---------- ---------- ----------
    CA_G.sIIb  272  G-AACGTTTG GAATGTACTC TATTT----- ---------- ----------
    NA_G.sIIb  272  G-AACGTTTG GAATGTACTC TATTT----- ---------- ----------
     NA_G.cal  270  ATGATTTGTT TGGAATACGA CTCTTTC--- ---------- ----------
       C_O.uni 293  TTAGTCCAAC TCGGTCCAGC TC-------- ---------- ----------
     CS_O.uni  293  TTAGTCCAAC TCGGTCCAGC TC-------- ---------- ----------
     CA_O.uni  296  CGGTTCGCCA GG-------- ---------- ---------- ----------
   M_O.uni(V)  296  CGGTTCGCCA GG-------- ---------- ---------- ----------
      C_G.sac  280  GATGAGAAAC AACTCGATCC GCTCAGC--- ---------- ----------
     CS_G.sac  280  GATGAGAAAC AACTCGATCC GCTCAGC--- ---------- ----------
    C_G.rub,P  278  TCGGATGTGC GACTATCGGT CTAGGATCGC ATCTCA-GCC CCTGG-----
   NA_G.rub,P  278  TCGGATGTGC GACTATCGGT CTAGGATCGC ATCTCA-GCC CCTGG-----
```

```
NA_G.rub,WI    279  TCGGATGTGC GACTATTGGT CTAGGATCAC ATGACTTGCC CCTGG-----
CS_G.rub,WI    279  TCGGATGTGC GACTATTGGT CTAGGATCAC ATGACTTGCC CCTGG-----
 C_G.rub(P)    283  TCGGATGTAC GACTATCGGT CTAGGATCAC ATGACTTGCC CCTGG-----
NA_G.rub,WII   275  CTTATGACCA TCACTCTAGA CGTTTATGTT TTGGTGTTGG CCCTTGA---
 C_G.rub,WII   275  CTTATGACCA TCACTCTAGA CGTTTATGTT TTGGTGTTGG CCCTTGA---
 C_G.con(P)    288  GCGACTATCA CTCTTAGGCA ATTATGATTT GCTTGTAGCC CCTGG-----
   CS_G.con    287  GCGACTATCA CTCTTAGGCA ACTATGATCT GCTTGTAGCC CCTGG-----
NA_N.pac(D)    325  TACATGGTGT AAATGGATTT GTT------- ---------- ----------
AA_N.pac(D)    325  TACATGGTGT AAATGGATTT GTT------- ---------- ----------
    C_N.dut    306  AGAGTTTAAA CATTGCGCAT GCTG------ ---------- ----------
CS_G.bul,Ia    282  ---------- ---------- ---------- ---------- ----------
CA_G.bul,IId   288  GAC------- ---------- ---------- ---------- ----------
 M_G.bul(V)    284  ---------- ---------- ---------- ---------- ----------
Na_G.bul,Ib    284  ---------- ---------- ---------- ---------- ----------
  NA_G.falc    256  ---------- ---------- ---------- ---------- ----------
  CS_G.falc    257  ---------- ---------- ---------- ---------- ----------
Na_G.bul,IIb   286  GAC------- ---------- ---------- ---------- ----------
 AA_G.bul,IIb  286  GAC------- ---------- ---------- ---------- ----------
Na-G.bul,IIa   286  GAC------- ---------- ---------- ---------- ----------
 AA_G.bul,IIa  286  GAC------- ---------- ---------- ---------- ----------
 AA_G.bul,IIc  289  GAC------- ---------- ---------- ---------- ----------
NA-T.qui,IIa   311  TAATTGCCAC TACGTCTGGT CCTTCGGGCA GAATAGTCCT TATGCTTGTC
 AA_T.qui,IIa  311  TAATTGCCAC TACGTCTGGT CCTTCGGGCA GAATAGTCCT TATGCTTGTC
NA_T.qui,IIb   311  TATTGCCACT ACGTCTGGTC CTTTGGGGGC AGAATAGTGT TGTATATTTT
 AA_T.qui,IIb  311  TATTGCCACT ACGTCTGGTC CTTTGGGCAG AATAGTGTTG TATATTTTGT
 CS_T.qui,I    369  CGGAAAGGTC ACATATAAAG TACGGTCTGC TAAAGATATG AATATCTTGC
   CS_G.glu    286  TCGCACGCTT TTTACA---- ---------- ---------- ----------
   NA_G.uvu    285  ---------- ---------- ---------- ---------- ----------
  A.becca_P    289  ACACGCGTGC CGCGCA---- ---------- ---------- ----------
    Troch_P    308  TTACTTT-GC TCGCACAA-- ---------- ---------- ----------
   H.germ_P    281  ---------- ---------- ---------- ---------- ----------
    Textu_P    300  TTGCTTTTGC TCACACAA-- ---------- ---------- ----------
    Biger_P    301  TTGCTTTTGC TCGCACAA-- ---------- ---------- ----------
    Boliv_P    284  CTCGCTGCGC T--------- ---------- ---------- ----------
   G.oper_P    284  CTTTTCATCG TTATCACATC ATACAAC--- ---------- ----------
  A.trian_P    302  TGANTTAATT TTCGTATATT ATATTATATA ATTATGTGAA TTTTTTGAGC
  As.rara_P    311  TATTATTATG TGCCAATATT ATTTTATTTT TTTTTAAAGG ----------
   A.angu_P    283  TAATAT---- ---------- ---------- ---------- ----------
   E.acul_P    228  ---------- ---------- ---------- ---------- ----------
    Quin_P     287  TTGATAATAT TA-------- ---------- ---------- ----------
  M.secan_P    286  TAATTTATAT TA-------- ---------- ---------- ----------
   P.pert_P    285  TTAAGTGAAC ATATTTTATT ATACATATAT TATTATATAT ATTTATTATA
 Allogrom_P    311  GCACGGTATT CTTTTAA--- ---------- ---------- ----------
      Align    170  ---------- ---------- ---------- ---------- ----------

       C_G.sI  330  ---------- ---------- ---------- ---------- ----------
      NA_G.sI  330  ---------- ---------- ---------- ---------- ----------
     C_G.sI(V) 330  ---------- ---------- ---------- ---------- ----------
     C_G.sIIa  291  ---------- ---------- ---------- ---------- ----------
     NA_G.sIIa 291  ---------- ---------- ---------- ---------- ----------
     CS_G.sIIa 290  ---------- ---------- ---------- ---------- ----------
     CA_G.sIIa 291  ---------- ---------- ---------- ---------- ----------
     CA_G.sIIb 296  ---------- ---------- ---------- ---------- ----------
     NA_G.sIIb 296  ---------- ---------- ---------- ---------- ----------
      NA_G.cal 297  ---------- ---------- ---------- ---------- ----------
       C_O.uni 315  ---------- ---------- ---------- ---------- ----------
      CS_O.uni 315  ---------- ---------- ---------- ---------- ----------
      CA_O.uni 308  ---------- ---------- ---------- ---------- ----------
     M_O.uni(V) 308  ---------- ---------- ---------- ---------- ----------
       C_G.sac 307  ---------- ---------- ---------- ---------- ----------
      CS_G.sac 307  ---------- ---------- ---------- ---------- ----------
     C_G.rub,P 322  ---------- ---------- ---------- ---------- ----------
    NA_G.rub,P 322  ---------- ---------- ---------- ---------- ----------
    NA_G.rub,WI 324  ---------- ---------- ---------- ---------- ----------
    CS_G.rub,WI 324  ---------- ---------- ---------- ---------- ----------
     C_G.rub(P) 328  ---------- ---------- ---------- ---------- ----------
   NA_G.rub,WII 322  ---------- ---------- ---------- ---------- ----------
    C_G.rub,WII 322  ---------- ---------- ---------- ---------- ----------
     C_G.con(P) 333  ---------- ---------- ---------- ---------- ----------
       CS_G.con 332  ---------- ---------- ---------- ---------- ----------
    NA_N.pac(D) 348  ---------- ---------- ---------- ---------- ----------
    AA_N.pac(D) 348  ---------- ---------- ---------- ---------- ----------
       C_N.dut  330  ---------- ---------- ---------- ---------- ----------
    CS_G.bul,Ia 282  ---------- ---------- ---------- ---------- ----------
   CA_G.bul,IId 291  ---------- ---------- ---------- ---------- ----------
    M_G.bul(V)  284  ---------- ---------- ---------- ---------- ----------
   Na_G.bul,Ib  284  ---------- ---------- ---------- ---------- ----------
```

```
   NA_G.falc   256 ---------- ---------- ---------- ---------- ----------
   CS_G.falc   257 ---------- ---------- ---------- ---------- ----------
Na_G.bul,IIb   289 ---------- ---------- ---------- ---------- ----------
  AA_G.bul,IIb 289 ---------- ---------- ---------- ---------- ----------
Na-G.bul,IIa   289 ---------- ---------- ---------- ---------- ----------
  AA_G.bul,IIa 289 ---------- ---------- ---------- ---------- ----------
  AA_G.bul,IIc 292 ---------- ---------- ---------- ---------- ----------
NA-T.qui,IIa   361 TACACAAAAT TCTCCACAAT GTTGTAGCCA TACTTGATTG TATGCGCTAT
  AA_T.qui,IIa 361 TACACAAAAT TCTCCACAAT GTTGTAGCCA TACTTGATTG TATGCGCTAT
NA_T.qui,IIb   361 GTCTTGTCAA CACAAAATTC TTCACAATAT TGTAGCCGTA CTTGATTGTA
  AA_T.qui,IIb 361 CTTGTCAACA CAAAATTCTT CACAATATTG TAGCCGTACT TGATTGTATG
  CS_T.qui,I   419 CCTAAAGATT G--------- ---------- ---------- ----------
    CS_G.glu   302 ---------- ---------- ---------- ---------- ----------
    NA_G.uvu   285 ---------- ---------- ---------- ---------- ----------
   A.becca_P   305 ---------- ---------- ---------- ---------- ----------
     Troch_P   325 ---------- ---------- ---------- ---------- ----------
    H.germ_P   281 ---------- ---------- ---------- ---------- ----------
     Textu_P   318 ---------- ---------- ---------- ---------- ----------
     Biger_P   319 ---------- ---------- ---------- ---------- ----------
     Boliv_P   295 ---------- ---------- ---------- ---------- ----------
    G.oper_P   311 ---------- ---------- ---------- ---------- ----------
   A.trian_P   352 TAAGGATTTG ATTCAAAGTA AAAG------ ---------- ----------
   As.rara_P   351 TAAGGATTTG ATTCAAAGTA AAAC------ ---------- ----------
    A.angu_P   289 ---------- ---------- ---------- ---------- ----------
    E.acul_P   228 ---------- ---------- ---------- ---------- ----------
     Quin_P    299 ---------- ---------- ---------- ---------- ----------
   M.secan_P   298 ---------- ---------- ---------- ---------- ----------
    P.pert_P   335 ---------- ---------- ---------- ---------- ----------
  Allogrom_P   328 ---------- ---------- ---------- ---------- ----------
       Align   170 ---------- ---------- ---------- ---------- ----------

      C_G.sI   330 ---------- ---------- ---------- ---------- ----------
     NA_G.sI   330 ---------- ---------- ---------- ---------- ----------
    C_G.sI(V)  330 ---------- ---------- ---------- ---------- ----------
     C_G.sIIa  291 ---------- ---------- ---------- ---------- ----------
    NA_G.sIIa  291 ---------- ---------- ---------- ---------- ----------
    CS_G.sIIa  290 ---------- ---------- ---------- ---------- ----------
    CA_G.sIIa  291 ---------- ---------- ---------- ---------- ----------
    CA_G.sIIb  296 ---------- ---------- ---------- ---------- ----------
    NA_G.sIIb  296 ---------- ---------- ---------- ---------- ----------
     NA_G.cal  297 ---------- ---------- ---------- ---------- ----------
      C_O.uni  315 ---------- ---------- ---------- ---------- ----------
     CS_O.uni  315 ---------- ---------- ---------- ---------- ----------
     CA_O.uni  308 ---------- ---------- ---------- ---------- ----------
   M_O.uni(V)  308 ---------- ---------- ---------- ---------- ----------
      C_G.sac  307 ---------- ---------- ---------- ---------- ----------
     CS_G.sac  307 ---------- ---------- ---------- ---------- ----------
    C_G.rub,P  322 ---------- ---------- ---------- ---------- ----------
   NA_G.rub,P  322 ---------- ---------- ---------- ---------- ----------
  NA_G.rub,WI  324 ---------- ---------- ---------- ---------- ----------
  CS_G.rub,WI  324 ---------- ---------- ---------- ---------- ----------
    C_G.rub(P) 328 ---------- ---------- ---------- ---------- ----------
  NA_G.rub,WII 322 ---------- ---------- ---------- ---------- ----------
   C_G.rub,WII 322 ---------- ---------- ---------- ---------- ----------
    C_G.con(P) 333 ---------- ---------- ---------- ---------- ----------
     CS_G.con  332 ---------- ---------- ---------- ---------- ----------
   NA_N.pac(D) 348 ---------- ---------- ---------- ---------- ----------
   AA_N.pac(D) 348 ---------- ---------- ---------- ---------- ----------
      C_N.dut  330 ---------- ---------- ---------- ---------- ----------
   CS_G.bul,Ia 282 ---------- ---------- ---------- ---------- ----------
  CA_G.bul,IId 291 ---------- ---------- ---------- ---------- ----------
    M_G.bul(V) 284 ---------- ---------- ---------- ---------- ----------
    Na_G.bul,Ib 284 ---------- ---------- ---------- ---------- ----------
     NA_G.falc 256 ---------- ---------- ---------- ---------- ----------
     CS_G.falc 257 ---------- ---------- ---------- ---------- ----------
  Na_G.bul,IIb 289 ---------- ---------- ---------- ---------- ----------
    AA_G.bul,IIb 289 ---------- ---------- ---------- ---------- ----------
  Na-G.bul,IIa 289 ---------- ---------- ---------- ---------- ----------
    AA_G.bul,IIa 289 ---------- ---------- ---------- ---------- ----------
    AA_G.bul,IIc 292 ---------- ---------- ---------- ---------- ----------
  NA-T.qui,IIa 411 TAATGTCTAT GTGTGACACA TTCGTGGTTC AAGACCAGTT CGGCTGCTTG
    AA_T.qui,IIa 411 TAATGTCTAT GTGTGACACA TTCGTGGTTC AAGACCAGTT CGGCTGCTTG
  NA_T.qui,IIb 411 TGCGCTATTC ATGTATATGT GGTTAGTCAA CACCAGTTCG GCTGCTTGCA
    AA_T.qui,IIb 411 CGCTATTCAT GTATATGTGG TTAGTCAACA CCAGTTCGGC TGCTTGCACA
    CS_T.qui,I   430 ---------- ---------- ---------- ---------- ----------
      CS_G.glu   302 ---------- ---------- ---------- ---------- ----------
      NA_G.uvu   285 ---------- ---------- ---------- ---------- ----------
```

```
A.becca_P   305 ---------- ---------- ---------- ---------- ----------
Troch_P     325 ---------- ---------- ---------- ---------- ----------
H.germ_P    281 ---------- ---------- ---------- ---------- ----------
Textu_P     318 ---------- ---------- ---------- ---------- ----------
Biger_P     319 ---------- ---------- ---------- ---------- ----------
Boliv_P     295 ---------- ---------- ---------- ---------- ----------
G.oper_P    311 ---------- ---------- ---------- ---------- ----------
A.trian_P   376 ---------- ---------- ---------- ---------- ----------
As.rara_P   375 ---------- ---------- ---------- ---------- ----------
A.angu_P    289 ---------- ---------- ---------- ---------- ----------
E.acul_P    228 ---------- ---------- ---------- ---------- ----------
Quin_P      299 ---------- ---------- ---------- ---------- ----------
M.secan_P   298 ---------- ---------- ---------- ---------- ----------
P.pert_P    335 ---------- ---------- ---------- ---------- ----------
Allogrom_P  328 ---------- ---------- ---------- ---------- ----------
Align       170 ---------- ---------- ---------- ---------- ----------

C_G.sI      330 ---------- ---------- ---------- ---------- ----------
NA_G.sI     330 ---------- ---------- ---------- ---------- ----------
C_G.sI(V)   330 ---------- ---------- ---------- ---------- ----------
C_G.sIIa    291 ---------- ---------- ---------- ---------- ----------
NA_G.sIIa   291 ---------- ---------- ---------- ---------- ----------
CS_G.sIIa   290 ---------- ---------- ---------- ---------- ----------
CA_G.sIIa   291 ---------- ---------- ---------- ---------- ----------
CA_G.sIIb   296 ---------- ---------- ---------- ---------- ----------
NA_G.sIIb   296 ---------- ---------- ---------- ---------- ----------
NA_G.cal    297 ---------- ---------- ---------- ---------- ----------
C_O.uni     315 ---------- ---------- ---------- ---------- ----------
CS_O.uni    315 ---------- ---------- ---------- ---------- ----------
CA_O.uni    308 ---------- ---------- ---------- ---------- ----------
M_O.uni(V)  308 ---------- ---------- ---------- ---------- ----------
C_G.sac     307 ---------- ---------- ---------- ---------- ----------
CS_G.sac    307 ---------- ---------- ---------- ---------- ----------
C_G.rub,P   322 ---------- ---------- ---------- ---------- ----------
NA_G.rub,P  322 ---------- ---------- ---------- ---------- ----------
NA_G.rub,WI 324 ---------- ---------- ---------- ---------- ----------
CS_G.rub,WI 324 ---------- ---------- ---------- ---------- ----------
C_G.rub(P)  328 ---------- ---------- ---------- ---------- ----------
NA_G.rub,WII 322 ---------- ---------- ---------- ---------- ----------
C_G.rub,WII 322 ---------- ---------- ---------- ---------- ----------
C_G.con(P)  333 ---------- ---------- ---------- ---------- ----------
CS_G.con    332 ---------- ---------- ---------- ---------- ----------
NA_N.pac(D) 348 ---------- ---------- ---------- ---------- ----------
AA_N.pac(D) 348 ---------- ---------- ---------- ---------- ----------
C_N.dut     330 ---------- ---------- ---------- ---------- ----------
CS_G.bul,Ia 282 ---------- ---------- ---------- ---------- ----------
CA_G.bul,IId 291 ---------- ---------- ---------- ---------- ----------
M_G.bul(V)  284 ---------- ---------- ---------- ---------- ----------
Na_G.bul,Ib 284 ---------- ---------- ---------- ---------- ----------
NA_G.falc   256 ---------- ---------- ---------- ---------- ----------
CS_G.falc   257 ---------- ---------- ---------- ---------- ----------
Na_G.bul,IIb 289 ---------- ---------- ---------- ---------- ----------
AA_G.bul,IIb 289 ---------- ---------- ---------- ---------- ----------
Na-G.bul,IIa 289 ---------- ---------- ---------- ---------- ----------
AA_G.bul,IIa 289 ---------- ---------- ---------- ---------- ----------
AA_G.bul,IIc 292 ---------- ---------- ---------- ---------- ----------
NA-T.qui,IIa 461 CACACATTGC CATCTCAACA TGATTGGTCA GGCTGCTAGA GACCTTGTGT
AA_T.qui,IIa 461 CACACATTGC CATCTCAACA TGATTGGTCA GGCTGCTAGA GACCTTGTGT
NA_T.qui,IIb 461 CACATCGCCA TCTCAACATG ATCAGTCAGG CTGCTAGAGA CCTTGTGTCT
AA_T.qui,IIb 461 CATCGCCATC TCAACATGAT CAGTCAGGCT GCTAGAGACC TTGTGTCTTG
CS_T.qui,I  430 ---------- ---------- ---------- ---------- ----------
CS_G.glu    302 ---------- ---------- ---------- ---------- ----------
NA_G.uvu    285 ---------- ---------- ---------- ---------- ----------
A.becca_P   305 ---------- ---------- ---------- ---------- ----------
Troch_P     325 ---------- ---------- ---------- ---------- ----------
H.germ_P    281 ---------- ---------- ---------- ---------- ----------
Textu_P     318 ---------- ---------- ---------- ---------- ----------
Biger_P     319 ---------- ---------- ---------- ---------- ----------
Boliv_P     295 ---------- ---------- ---------- ---------- ----------
G.oper_P    311 ---------- ---------- ---------- ---------- ----------
A.trian_P   376 ---------- ---------- ---------- ---------- ----------
As.rara_P   375 ---------- ---------- ---------- ---------- ----------
A.angu_P    289 ---------- ---------- ---------- ---------- ----------
E.acul_P    228 ---------- ---------- ---------- ---------- ----------
Quin_P      299 ---------- ---------- ---------- ---------- ----------
M.secan_P   298 ---------- ---------- ---------- ---------- ----------
P.pert_P    335 ---------- ---------- ---------- ---------- ----------
```

```
Allogrom_P   328  ---------- ---------- ---------- ---------- ----------
     Align   170  ---------- ---------- ---------- ---------- ----------

     C_G.sI  330  ---------- ---------- -------GAC TCAATTGAAC GCAACGGACG
    NA_G.sI  330  ---------- ---------- -------GAC TCAATTGAAC GCAACGGACG
  C_G.sI(V)  330  ---------- ---------- -------GAC TCAATTGAAC GCAACGGACG
  C_G.sIIaa  291  ---------- ---------- -------GGC TCAATTGAAC GCAACGGACG
   NA_G.sIIa 291  ---------- ---------- -------GGC TCAATTGAAC GCAACGGACG
   CS_G.sIIa 290  ---------- ---------- -------GGC TCAATTGAAC GCAACGGACG
   CA_G.sIIb 291  ---------- ---------- -------GGC TCAATTGAAC GCAACGGACG
   CA_G.sIIb 296  ---------- ---------- -------GGC TCAATTGAAC GCAACGGACG
    NA_G.sII 296  ---------- ---------- -------GGC TCAATTGAAC GCAACGGACG
    NA_G.cal 297  ---------- ---------- -------GGC TCAATTGAAC GCAACGGACG
     C_O.uni 315  ---------- ---------- -------GAT TCTTTCCAAT GCAACGAGCG
    CS_O.uni 315  ---------- ---------- -------GAT TCTTTCCAAT GCAACGAGCG
    CA_O.uni 308  ---------- ---------- -------GTC TCTTTTGAAC GCAACGGACG
  M_O.uni(V) 308  ---------- ---------- -------GTC TCTTTTGAAC GCAACGGACG
     C_G.sac 307  ---------- ---------- -------GTT CTATTGGAAT GCAACGGACG
    CS_G.sac 307  ---------- ---------- -------GTT CTATTGGAAT GCAACGGACG
   C_G.rub,P 322  ---------- ---------- -------GAC TCTTTTGAAC GCAACGGACG
  NA_G.rub,P 322  ---------- ---------- -------GAC TCTTTTGAAC GCAACGGACG
 NA_G.rub,WI 324  ---------- ---------- -------GAC TCTTTTGAAC GCAACGGACG
 CS_G.rub,WI 324  ---------- ---------- -------GAC TCTTTTGAAC GCAACGGACG
   C_G.rub(P) 328  ---------- ---------- -------GAC TCTTTTGAAC GCAACGGACG
 NA_G.rub,WII 322  ---------- ---------- -------TGC TCATTTGAAC GCAACGGACG
  C_G.rub,WII 322  ---------- ---------- -------TGC TCATTTGAAC GCAACGGACG
   C_G.con(P) 333  ---------- ---------- -------TAC TCTTTTGAAC GCAACGGACG
    CS_G.con  332  ---------- ---------- -------TAC TCTTTTGAAC GCAACGGACG
  NA_N.pac(D) 348  ---------- ---------- -----TTGGG TACCCAGAAA GCAACGAACG
  AA_N.pac(D) 348  ---------- ---------- -----TTGGG TACCCAGAAA GCAACGAACG
     C_N.dut  330  ---------- ---------- -----TTGGG --TCCTGAAA GCAACGAACG
  CS_G.bul,Ia 282  ---------- ---------- -------ATT TGAACAGTAC GCAACGGACG
 CA_G.bul,IId 291  ---------- ---------- -------CTC TGAACAGTAC GCAACGAACG
   M_G.bul(V) 284  ---------- ---------- -------ACT TGAACAGTAC GCAACGGACG
  Na_G.bul,Ib 284  ---------- ---------- -------ACT TGAACAGTAC GCAACGGACG
    NA_G.falc 256  ---------- ---------- -------CTT TTGTGTGAAC GCAACGGACG
    CS_G.falc 257  ---------- ---------- -------CTT TTGTGTGAAC GCAACGGACG
 Na_G.bul,IIb 289  ---------- ---------- -------CTC TGAACAGTAC GCAACGAACG
  AA_G.bul,IIb 289  ---------- ---------- -------CTC TGAACAGTAC GCAACGAACG
 Na-G.bul,IIa 289  ---------- ---------- -------CTC TGAACAGTAC GCAACGAACG
  AA_G.bul,IIa 289  ---------- ---------- -------CTC TGAACAGTAC GCAACGAACG
  AA_G.bul,IIc 292  ---------- ---------- -------CTC TGAACAGTAC GCAACGAACG
 NA-T.qui,IIa 511  CTTGCCCCAA AGGTTG---- -----ATTAA CACCGTGAGT GCAACGAGTG
  AA_T.qui,IIa 511  CTTGCCCCAA AGGTTG---- -----ATTAA CACCGTGAGT GCAACGAGTG
 NA_T.qui,IIb 511  TGCCCCAAAG GTTG------ -----ATTAA CACCGTGAGT GCAACGAGTG
  AA_T.qui,IIb 511  CCCCAAAGGT TG-------- -----ATTAA CACCGTGAGT GCAACGAGTG
   CS_T.qui,I 430  ---------- ---------- -----ATTAA CACCGTGAGT GCAACGAGTG
     CS_G.glu 302  ---------- ---------- -------TTA GGTCCTGAAA GCAACGAACG
     NA_G.uvu 285  ---------- ---------- -------TAT CCTTTCGAAT GCAACGAACG
    A.becca_P 305  ---------- ---------- -------CTG GTCTCAGATA GCAACGAACG
      Troch_P 325  ---------- ---------- -------TTA AGTCCTGAAA GCAACGAACG
      H.germ_P 281  ---------- ---------- -------TTG TGCTTTGAAA GCAACGAACG
      Textu_P 318  ---------- ---------- -------TTA GGTCCTGAAA GCAACGAACG
      Biger_P 319  ---------- ---------- -------TTA GGTCCTGAAA GCAACGAACG
      Boliv_P 295  ---------- ---------- -------TTA GATCCTGAAA GCAACGAACG
     G.oper_P 311  ---------- ---------- -------GTT GATTCTGAAA GCAACGAACG
    A.trian_P 376  ---------- ---------- -------CTG GAGCCTGAAG GCAACGAACG
    As.rara_P 375  ---------- ---------- -------GTG GAGCCTGAAG GCAACGAACG
     A.angu_P 289  ---------- ---------- -------ATA ATAAATGAAT GCAACGAACG
     E.acul_P 228  ---------- ---------- -------GTA CACTTTGAAA GCAACGAACG
       Quin_P 299  ---------- ---------- -------TAT TTAAATGAAT GCAACGAACG
    M.secan_P 298  ---------- ---------- -------TTA ATAATTGAAT GCAACGAACG
     P.pert_P 335  ---------- ---------- -------TAT TAAAATGAAT GCAACGAACG
   Allogrom_P 328  ---------- ---------- -------ATA TACTCTGAAG GCAACGAACG
        Align 170  ---------- ---------- ---------- ---------- mmmmmmmmmmmm

     C_G.sI  353  TGATTGC-AA ----GTCCTT GT--TG---- ---------- ---AACA-AA
    NA_G.sI  353  TGATTGC-AA ----GTCCTT GT--TG---- ---------- ---AACA-AA
  C_G.sI(V)  353  TGATTGC-AA ----GTCCTT GT--TG---- ---------- ---AACA-AA
   C_G.sIIa  314  TGATCGC-GA ----GTCCTT GT--TG---- ---------- ---AACT-TC
   NA_G.sIIa 314  TGATCGC-GA ----GTCCTT GT--TG---- ---------- ---AACT-TC
   CS_G.sIIa 313  TGATCGC-GA ----GTCCTT GT--TG---- ---------- ---AACT-TC
   CA_G.sIIa 314  TGATCGC-GA ----GTCCTT GT--TG---- ---------- ---AACT-TC
   CA_G.sIIb 319  TGATCGC-AA ----GTCCTT GT--TG---- ---------- ---AACT-TC
   NA_G.sIIb 319  TGATCGC-AA ----GTCCTT GT--TG---- ---------- ---AACT-TC
    NA_G.cal 320  TGATTGC-AA ----GTCCTT GT--TG---- ---------- ---AACT-TC
     C_O.uni 338  TGATCGC-GA ----CCCCTT GT--TG---- ---------- ---AGCT-TA
```

```
    CS_O.uni   338  TGATCGC-GA ----CCCCTT GT--TG---- ---------- ---AGCT-TA
    CA_O.uni   331  TGATCGC-AA ----CCCTTT GT--TG---- ---------- ---AGTGATG
  M_O.uni(V)   331  TGATCGC-AA ----CCCTTT GT--TG---- ---------- ---AGTGATG
     C_G.sac   330  TGATTGC-AA ----GCTTTT GTTTTG---- ---------- ---AGTT-GC
    CS_G.sac   330  TGATTGC-AA ----GCTTTT GTTTTG---- ---------- ---AGTT-GC
   C_G.rub,P   345  TGATTGC-AA ----CCCCTT GT--TG---- ---------- ---AGAT-TA
  NA_G.rub,P   345  TGATTGC-AA ----CCCCTT GT--TG---- ---------- ---AGAT-TA
 NA_G.rub,WI   347  TGATTGC-AA ----CCCCTT GT--TG---- ---------- ---AGAT-TA
 CS_G.rub,WI   347  TGATTGC-AA ----CCCCTT GT--TG---- ---------- ---AGAT-TA
   C_G.rub(P)  351  TGATTGC-AA ----CCCCTT GT--TG---- ---------- ---AGAT-TA
NA_G.rub,WII   345  TGATTGC-AA ----CCCCTT GT--TG---- ---------- ---AGATTTA
 C_G.rub,WII   345  TGATTGC-AA ----CCCCTT GT--TG---- ---------- ---AGATTTA
   C_G.con(P)  356  TGATTGC-AA ----CCCCTT GT--TG---- ---------- ---AGTTTGC
    CS_G.con   355  TGATTGC-AA ----CCCCTT GT--TG---- ---------- ---AGTTTGC
  NA_N.pac(D)  373  TGACCGC-AA ----CGTCTT GT--TG---- ---------- ---TCTCTCT
  AA_N.pac(D)  373  TGACCGC-AA ----CGTCTT GT--TG---- ---------- ---TCTCTCT
     C_N.dut   353  TGACCGC-AA ----CGTCTT AT--TG---- ---------- ---CCTT-TA
  CS_G.bul,Ia  305  CGATCGT-AA ----TCTCTT GT--TA---- ---------- ---AGTGGCC
 CA_G.bul,IId  314  CGATCGT-AA ----TCCCTT GT--TG---- ---------- ---AGTGGCC
   M_G.bul(V)  307  TGATCGT-AA ----TCTCTT GT--TA---- ---------- ---AATGGTT
  Na_G.bul,Ib  307  TGATCGT-AA ----TCTCTT GT--TA---- ---------- ---AATGGTT
    NA_G.falc  279  TGATTGC-AA ----CCCCTT GT--TG---- ---------- ---ATTGGCC
    CS_G.falc  280  TGATTGC-AA ----CCCCTT GT--TG---- ---------- ---ATTGGCC
 Na_G.bul,IIb  312  CGATCGT-AA ----TCCCTT GT--TG---- ---------- ---AGTGGCC
  AA_G.bul,IIb 312  CGATCGT-AA ----TCCCTT GT--TG---- ---------- ---AGTGGCC
 Na-G.bul,IIa  312  CGATCGT-AA ----TCCCTT GT--TG---- ---------- ---AGTGGCC
  AA_G.bul,IIa 312  CGATCGT-AA ----TCCCTT GT--TG---- ---------- ---AGTGGCC
  AA_G.bul,IIc 315  CGATCGT-AA ----TCCCTT GT--TG---- ---------- ---AGTGGCC
 NA-T.qui,IIa  552  AGATTGC-AA -----GTCTT TG--TT---- ---------- ---ATGTAGT
  AA_T.qui,IIa 552  AGATTGC-AA -----GTCTT TG--TT---- ---------- ---ATGTAGT
 NA_T.qui,IIb  550  AGATTGC-AA -----GCCTT TG--TT---- ---------- ---ATGTAGT
  AA_T.qui,IIb 548  AGATTGC-AA -----GCCTT TG--TT---- ---------- ---ATGTAGT
  CS_T.qui,I   455  AGATTGC-GA -----GTCTT TG--TT---- ---------- ---ATGTAGT
    CS_G.glu   325  TGACCGC-AA ----CCTCTT GT--TG---- ---------- ---CCTCATC
    NA_G.uvu   308  TGACCGC-AA ----CCTCTT GT--TG---- ---------- ---CCTTCGT
   A.becca_P   328  TGACCGT-AT ----TCTATT GT--TG---- ---------- ---CAGTGAA
     Troch_P   348  TGACCGC-AA ----CCTCTT GT--TG---- ---------- ---CCTTTAG
     H.germ_P  304  TGACCGC-AA ----CCTCTT GT--TG---- ---------- ---CCTGTAT
      Textu_P  341  TGACCGC-AA ----CCTCTT GT--TG---- ---------- ---CCTTTAT
      Biger_P  342  TGACCGC-AA ----CCTCTT GT--TG---- ---------- ---CCTTTAT
      Boliv_P  318  TGACCGC-AA ----CCTCTT GT--TG---- ---------- ---CTTTCAT
     G.oper_P  334  TGACCGC-AA ----CCTCTT GT--TG---- ---------- ---CCTGTAT
    A.trian_P  399  TGACCGC-AA ----CCTCTT GT--TG---- ---------- ---CCTCATT
    As.rara_P  398  TGACCGC-AA ----CCTCTT GT--TG---- ---------- ---CCTCATT
     A.angu_P  312  TGACCGT-AA ----CCTTTT AT--TG---- ---------- ---CTATAAT
     E.acul_P  251  TGACCGT-AT ----TCTTTT GT--TA---- ---------- ---TATATAC
       Quin_P  322  TGACTAT-AA ----CCTTTT AT--TG---- ---------- ---CAATATT
    M.secan_P  321  TGACTAT-AA ----CCCTTT AT--TG---- ---------- ---CTTTATA
     P.pert_P  358  TGACCGT-AA ----CCTTTT AT--TG---- ---------- ---CTATAAA
   Allogrom_P  351  TGACCGC-AA ----CATCTT GT--TG---- ---------- ---CATAATC
        Align  180  mmmmmmmm-mm -----mmmmm mm--mm---- ---------- ----------

     C_G.sI    378  TT-A-TATAT AT--A-CTAC TTCATCATTA ----GTATAT ATAG--TTC-
    NA_G.sI    378  TT-A-TATAT AT--A-CTAC TTCATCATTA ----GTATAT ATAG--TTC-
   C_G.sI(V)   378  TT-A-TATAT AT--A-CTAC TTCATCATTA ----GTATAT ATAG--TTC-
   C_G.sIIa    339  AG---TATAT AT-TGACTAT CCCCGTAGGA AGGGATTGGC AATAATTGAA
  NA_G.sIIa    339  AG---TATAT AT-TGACTAT CCCCGTAGGA AGGGATTGGC AATAATTGAA
  CS_G.sIIa    338  AG---TATAT AT-TGACTTT CCCCGTAGGA AGGGATTGGC AATAATTGAA
  CA_G.sIIa    339  AG---TATAT AT-TGACTAT CCCCGTAGGA AGGGATTGGC AATAATTGAA
  CA_G.sIIb    344  AG---TATAT AT-TGACTTT CCCCGTAGGA AGGGATTGCC AATAATTGAA
  NA_G.sIIb    344  AG---TATAT AT-TGACTTT CCCCGTAGGA AGGGATTGCC AATAATTGAA
   NA_G.cal    345  -----TATAC ATATCATTTA CATTCTTAAC TGAGTGTATC TAAGATAATA
    C_O.uni    363  TTACACATAA GCTCTATGTC GACGGAACGC TTTTGCGTTA CGGA------
   CS_O.uni    363  TTACACATAA GCTCTATGTC GACGGAACGC TTTTGCGTTA CGGA------
   CA_O.uni    357  TCAGCACTCT ACGTCTATCC TAGCTCGTTG ATCGCAAGTG AGAGTTGGGT
  M_O.uni(V)   357  TCAGCACCCT ACGTCTATCC TAGCTCGTTG ATCGCAAGTG AGAGTTGGGT
     C_G.sac   357  GG---TTTTA CTGAAAGACT CTACAATCCG CAAACTTCAA TGATCTGAAG
    CS_G.sac   357  GG---TTTTA CTGAAAGACT CTACAATCCG CAAACTTCAA TGATCTGAAG
   C_G.rub,P   370  GAGG------ ---AGTCTCG ATATTCATTG GCC--AAACT TACCGCTTCG
  NA_G.rub,P   370  GAGG------ ---AGTCTCG ATATTCATTG GCC--AAACT TACCGCTTCG
 NA_G.rub,WI   372  GAGAT----- ---AGTCTCG ATATTCACTG GAC--AAACT GGCTGCTTTA
 CS_G.rub,WI   372  GAGAT----- ---AGTCTCG ATATTCACTG GAC--AAACT GGCTGCTTTA
   C_G.rub(P)  376  GAGAT----- ---AGTCTCG ATATTCACTG GCC--AAACT GGCTGCTTTA
NA_G.rub,WII   371  TCCGATCTCT ACGTTCACTG AACCACACTC TTTAGCGAGA GCAGTGGGGA
 C_G.rub,WII   371  TCCGATCTCT ACGTTCACTG AACCACACTC TTTAGCGAGA GCAGTGGGGA
   C_G.con(P)  382  TTCTCTTAGG AACTCTACGT CTACTGATCC ACCTACCTTA GCGGGTAGGT
    CS_G.con   381  TTCTCTTAGG AACTCTACGT CTACTGATCC ACCTACCTTA GCGGGTAGGT
```

```
NA_N.pac(D)    399  TTGACAGTTA TGGGTTATCC CAGTCATGTG TTTATACTTT TATGTGTAAA
AA_N.pac(D)    399  TTGACAGTTA TGGGTTATCC CAGTCATGTG TTTATACTTT TATGTGTAAA
    C_N.dut    378  TCTTGCTATA TTCTAATTTA ATTAGAAATA GCTAACAGAG GCTAAT----
CS_G.bul,Ia    331  ATCCTGTGAG CCCCTGATTT AATGGCAGGC GGTATCATCT CAGCCACATT
CA_G.bul,IId   340  ATCCTGTAAG CTGCTGGATT AGGAACCCAG TGGTATTATC TCAGCCACAG
  M_G.bul(V)   333  ATCCTGTGAG CCAACCGTGA CTGATTACAG CACACCTGGT ATCATCCAGA
Na_G.bul,Ib    333  ATCCTGTGAG CCAACCGTGA CTGATTACAG CACACCTGGT ATCATCCAGA
   NA_G.falc   305  ATCACGTAAG CTCACTTGTA ATTTATTACA GTGTGGTATT ATCTTAGCCA
   CS_G.falc   306  ATCACGTAAG CTCACTTGTA ATTTATTACA GTGTGGTATT ATCTTAGCCA
Na_G.bul,IIb   338  ATCCTGTAAG CTGCTGGATT AGGAACCCAG TGGTATTATC TCAGCCACAG
  AA_G.bul,IIb 338  ATCCTGTAAG CTGCTGGATT AGGAACCCAG TGGTATTATC TCAGCCACAG
Na-G.bul,IIa   338  ATCCTGTAAG CTGCTGGATT AGGAACCCAG TGGTATTATC TCAGCCACAG
  AA_G.bul,IIa 338  ATCCTGTAAG CTGCTGGATT AGGAACCCAG TGGTATTATC TCAGCCACAG
  AA_G.bul,IIc 341  ATCCTGTAAG CTGCTGGATT TGAAACCCAG TGGTATTATC TCAGCCACAG
NA-T.qui,IIa   577  GAACAACATA TACCTACTAC TCTAGTAGTA GGATTCCAAC TACACAGTAT
  AA_T.qui,IIa 577  GAACAACATA TACCTACTAC TCTAGTAGTA GGATTCCAAC TACACAGTAT
NA_T.qui,IIb   575  GAACAACATA TACCTACTAC TACTACTATT TTGTAGTAAT GTAGGATTCC
  AA_T.qui,IIb 573  GAACAACATA TACCTACTAC TACTACTTTG TAGTAATGTA GGATTCCAAC
  CS_T.qui,I   480  GCTCAACATA CCTATGTGCA GTAGGATTTA ACTACACCAT GTAACCTTTT
    CS_G.glu   351  TCCCTACTTT TTTGAGCACT TCGGTGTTGA ATAATTTCAT TGAGGCTTT-
    NA_G.uvu   334  TTAACATCCT GCAGCTCGCT GTGGGTACTG AGGGCTTCAC TCGGTGACTT
   A.becca_P   354  TATGTGTGCC TTCGGGCGTA CGACCCACTG CTTAGTATAT GCACGCCTCG
     Troch_P   374  ATTTAACACT GTTTAGTCAT ATTTATTATG CATCTATCAG T--------A
     H.germ_P  330  ATATGTGTAT TTTATACACA CCACAGGCTA T--------- ----------
      Textu_P  367  CCCAAACGCC GTATATAATT TATTATTTCG GT-------- ---------A
      Biger_P  368  CCCAAACGTC GTTTAACGTT TACATTTAAT GTATCGTTGC GGC------A
      Boliv_P  344  ACCCAATGCG CTGTAATATC ACTCGTGATA TCTCAGCGCA TAAGAAGGCT
     G.oper_P  360  TACCACAACA GTCTGCACTC TGTNTTTCTG TATAAACAGG CCTTATATT-
    A.trian_P  425  CTTAATATGA ATGATATATT TATTTGTTTT ATTACATTTA AATGTATTAT
    As.rara_P  424  CTTAATATGA ATATTAATAT TTAAAATGAT TTTATTATTG TTTTATTTAT
     A.angu_P  338  AATTATTAAT ATAG------ ---------- ---------- ----------
     E.acul_P  277  TATA------ ---------- ---------- ---------- ----------
       Quin_P  348  TTTATATTAA TATAATTTTT AATTATATTA ATTATATTG- ----------
    M.secan_P  347  TTTATTAAAT TAAATAAAG- ---------- ---------- ----------
     P.pert_P  384  TAATATATAT TATTTATATT ATAATAG--- ---------- ----------
   Allogrom_P  377  TTATTTA--- ---------- ---------- ---------- ----------
        Align  198  ---------- ---------- ---------- ---------- ----------

       C_G.sI  416  --GCTTCTC- ---ATGTCGT AG--CAGTCA AACGGGCGGG CGTCTTAATT
      NA_G.sI  416  --GCTTCTC- ---ATGTCGT AG--CAGTCA AACGGGCGGG CGTCTTAATT
    C_G.sI(V)  416  --GCTTCTC- ---ATGTCGT AG--CAGTCA AACGGGCGGG CGTCTTAATT
     C_G.sIIa  385  TGTTCG--CT TTCTATTAGG TGTGTATATA TAGTCTCCAA ATTACATGTA
    NA_G.sIIa  385  TGTTCG--CT TTCTATTAGG TGTGTATATA TAGTCTCAAA ATTACATGTA
    CS_G.sIIa  384  TGTTCG--CT TTCTATTAGG TTTGTATGT- -AATCTCAAA ATTACATGTA
    CA_G.sIIa  385  TGTTCG--CT TTCTATTAGG TGTGTATATA TAGTCTCAAA ATTACATGTA
    CA_G.sIIb  390  TGTTCG--CT TTCTATT-GG TGTGTGATCT CAAAATTGCA TGTATTTGTG
    NA_G.sIIb  390  TGTTCG--CT TTCTATT-GG TGTGTGATCT CAAAATTGCA TGTATTTGTG
     NA_G.cal  390  GAGTTCG-CT TTCTATT-GG TGACGGGATC TCCAATCACT TTCGAGTGAG
       C_O.uni 407  ---------- ---------- ---------- ---------- ----------
      CS_O.uni 407  ---------- ---------- ---------- ---------- ----------
      CA_O.uni 407  G--------- ---------- ---------- ---------- ----------
    M_O.uni(V) 407  G--------- ---------- ---------- ---------- ----------
       C_G.sac 404  TAG---AAAA CCTAC----- ATGCT-ACAG TGAGATATTG TG--------
      CS_G.sac 404  TAG---AAAA CCTAC----- ATGCT-ACAG TGAGATATTG TG--------
     C_G.rub,P 409  AGCTTG--CT CGTGCAGTTT T-GTTAGCCG GTG------- ----------
    NA_G.rub,P 409  AGCTTG--CT CGTGCAGTTT T-GTTAGCCG GTG------- ----------
   NA_G.rub,WI 412  TGCAGT--CT TGTTGAAT-- ---------G GTG------- ----------
   CS_G.rub,WI 412  TGCAGT--CT TGTTGAAT-- ---------G GTG------- ----------
     C_G.rub(P) 416  TGCAGT--CT TGTTGAAT-- ---------G GTG------- ----------
  NA_G.rub,WII 421  TGTG------ ---------- ---------- ---------- ----------
   C_G.rub,WII 421  TGTG------ ---------- ---------- ---------- ----------
     C_G.con(P) 432  GGTGAGGTG- ---------- ---------- ---------- ----------
      CS_G.con  431  GGTGAGGTG- ---------- ---------- ---------- ----------
   NA_N.pac(D)  449  TACGTACGAC ACAGAGACTA GAT------- ---------- ----------
   AA_N.pac(D)  449  TACGTACGAC ACAGAGACTA GAT------- ---------- ----------
       C_N.dut  424  ---------- ---------- ---------- ---------- ----------
  CS_G.bul,Ia   381  TCCTCTGGTA GTAGTGGGCC AGAT------ ---------- ----------
 CA_G.bul,IId   390  ATTTTCTGGT TGTAGTGGGC CAG------- ---------- ----------
    M_G.bul(V)  383  CCATATTGGT TTAGTGGGTC GGGCTAGAT- ---------- ----------
  Na_G.bul,Ib   383  CCATATTGGT TTAGTGGGTC GGGCTAGAT- ---------- ----------
     NA_G.falc  355  TTCCCTTTTA GTTGTGGAGT GATTAACCAC TCTAAA---- ----------
     CS_G.falc  356  TTCCCTTTTA GTTGTGGCGT GATTAACCAC TCTAAA---- ----------
  Na_G.bul,IIb  388  ATTTTCTGGT TGTAGTGGGC CAG------- ---------- ----------
  AA_G.bul,IIb  388  ATTTTCTGGT TGTAGTGGGC CAG------- ---------- ----------
 Na-G.bul,IIa   388  ATTTTCTGGT TGTAATGGGC CAG------- ---------- ----------
  AA_G.bul,IIa  388  ATTTTCTGGT TGTAATGGGC CAG------- ---------- ----------
  AA_G.bul,IIc  391  ATTTTCTGGT TGTAATGGGC CAG------- ---------- ----------
```

```
NA-T.qui,IIa    627  AACCATTAAG GTGA------ ---------- ---------- ----------
  AA_T.qui,IIa  627  AACCATTAAG GTGA------ ---------- ---------- ----------
NA_T.qui,IIb    625  AACTACACAG AATAACCATT AAGGTGA--- ---------- ----------
  AA_T.qui,IIb  623  TACACACA-G AATAACCATT AAGGTGA--- ---------- ----------
  CS_T.qui,I    530  TCAAAGGTTG ---------- ---------- ---------- ----------
    CS_G.glu    400  ---------- ---------- ---------- ---------- ----------
    NA_G.uvu    384  TGGGCTGAGC AATCGGTCTT TTTGATCTG- ---------- ----------
   A.becca_P    404  GCGAGCAATA TAC------- ---------- ---------- ----------
     Troch_P    416  AAAAAAGGCT ---------- ---------- ---------- ----------
    H.germ_P    361  ---------- ---------- ---------- ---------- ----------
     Textu_P    400  AAAAAAGGCT ---------- ---------- ---------- ----------
     Biger_P    412  AAAAAAGGCT ---------- ---------- ---------- ----------
     Boliv_P    394  TATACATATA TCGCTACGGC GATGTAT--- ---------- ----------
    G.oper_P    409  ---------- ---------- ---------- ---------- ----------
   A.trian_P    475  TTATATGTGT TATGTATTTG ATAGTATATG AATGTTATGT GACATGTGTA
   As.rara_P    474  TCTTATTTAT ATGTGTTATG TATTTGATTG TATGTGAATG TATGTTACAT
    A.angu_P    352  ---------- ---------- ---------- ---------- ----------
    E.acul_P    281  ---------- ---------- ---------- ---------- ----------
       Quin_P   387  ---------- ---------- ---------- ---------- ----------
   M.secan_P    366  ---------- ---------- ---------- ---------- ----------
    P.pert_P    411  ---------- ---------- ---------- ---------- ----------
  Allogrom_P    384  ---------- ---------- ---------- ---------- ----------
       Align    198  ---------- ---------- ---------- ---------- ----------

      C_G.sI    458  GGCGCGGTCA CGAGG---CT TAATATGCTG CGGCAG---- ----------
     NA_G.sI    458  GGCGCGGTCA CGAGG---CT TAATATGCTG CGGCAG---- ----------
   C_G.sI(V)    458  GGCGCGGTCA CGAGG---CT TAATATGCTG CGGCAG---- ----------
    C_G.sIIa    433  TTTGTAATGG CGAGTAATAT TTA--CTATC CTGTA----- ----------
   NA_G.sIIa    433  TTTGTAATGG CGAGTAATAT TTATACTATC CTGTA----- ----------
   CS_G.sIIa    430  TTTGTAATGG CGAGTGAT-- -CA--CA--C CTATA----- ----------
   CA_G.sIIa    433  TTTGTAATGG CGAGTAATAT ATT--ACTAT CCTGTA---- ----------
   CA_G.sIIb    437  ATGGCGAGTG ATCTAATCCC TTA------- ---------- ----------
   NA_G.sIIb    437  ATGGCGAGTG ATCTAATCCC TTA------- ---------- ----------
    NA_G.cal    438  AGT--GAGTG ATCTAATCCT TTA------- ---------- ----------
      C_O.uni   407  ---------- ---------- ---------- ---------- ----------
     CS_O.uni   407  ---------- ---------- ---------- ---------- ----------
     CA_O.uni   408  ---------- ---------- ---------- ---------- ----------
   M_O.uni(V)   408  ---------- ---------- ---------- ---------- ----------
      C_G.sac   437  ---------- ---------- ---------- ---------- ----------
     CS_G.sac   437  ---------- ---------- ---------- ---------- ----------
    C_G.rub,P   439  ---------- ---------- ---------- ---------- ----------
   NA_G.rub,P   439  ---------- ---------- ---------- ---------- ----------
  NA_G.rub,WI   432  ---------- ---------- ---------- ---------- ----------
  CS_G.rub,WI   432  ---------- ---------- ---------- ---------- ----------
   C_G.rub(P)   436  ---------- ---------- ---------- ---------- ----------
 NA_G.rub,WII   425  ---------- ---------- ---------- ---------- ----------
  C_G.rub,WII   425  ---------- ---------- ---------- ---------- ----------
   C_G.con(P)   441  ---------- ---------- ---------- ---------- ----------
     CS_G.con   440  ---------- ---------- ---------- ---------- ----------
  NA_N.pac(D)   472  ---------- ---------- ---------- ---------- ----------
  AA_N.pac(D)   472  ---------- ---------- ---------- ---------- ----------
      C_N.dut   424  ---------- ---------- ---------- ---------- ----------
  CS_G.bul,Ia   405  ---------- ---------- ---------- ---------- ----------
 CA_G.bul,IId   413  ---------- ---------- ---------- ---------- ----------
   M_G.bul(V)   412  ---------- ---------- ---------- ---------- ----------
  Na_G.bul,Ib   412  ---------- ---------- ---------- ---------- ----------
    NA_G.falc   391  ---------- ---------- ---------- ---------- ----------
    CS_G.falc   392  ---------- ---------- ---------- ---------- ----------
 Na_G.bul,IIb   411  ---------- ---------- ---------- ---------- ----------
  AA_G.bul,IIb  411  ---------- ---------- ---------- ---------- ----------
 Na-G.bul,IIa   411  ---------- ---------- ---------- ---------- ----------
  AA_G.bul,IIa  411  ---------- ---------- ---------- ---------- ----------
  AA_G.bul,IIc  414  ---------- ---------- ---------- ---------- ----------
NA-T.qui,IIa    641  ---------- ---------- ---------- ---------- ----------
  AA_T.qui,IIa  641  ---------- ---------- ---------- ---------- ----------
NA_T.qui,IIb    652  ---------- ---------- ---------- ---------- ----------
  AA_T.qui,IIb  649  ---------- ---------- ---------- ---------- ----------
  CS_T.qui,I    540  ---------- ---------- ---------- ---------- ----------
    CS_G.glu    400  ---------- ---------- ---------- ---------- ----------
    NA_G.uvu    413  ---------- ---------- ---------- ---------- ----------
   A.becca_P    417  ---------- ---------- ---------- ---------- ----------
     Troch_P    426  ---------- ---------- ---------- ---------- ----------
    H.germ_P    361  ---------- ---------- ---------- ---------- ----------
     Textu_P    410  ---------- ---------- ---------- ---------- ----------
     Biger_P    422  ---------- ---------- ---------- ---------- ----------
     Boliv_P    421  ---------- ---------- ---------- ---------- ----------
    G.oper_P    409  ---------- ---------- ---------- ---------- ----------
```

```
A.trian_P   525  TAATATTATA TATATATATA TATTATATAT GANATGATTT GTATTGTTAG
As.rara_P   524  GTATTTAATC TGATGCATTC ATTGTGGTTA TATATATATA TGTATTTATA
A.angu_P    352  ---------- ---------- ---------- ---------- ----------
E.acul_P    281  ---------- ---------- ---------- ---------- ----------
Quin_P      387  ---------- ---------- ---------- ---------- ----------
M.secan_P   366  ---------- ---------- ---------- ---------- ----------
P.pert_P    411  ---------- ---------- ---------- ---------- ----------
Allogrom_P  384  ---------- ---------- ---------- ---------- ----------
Align       198  ---------- ---------- ---------- ---------- ----------

C_G.sI      491  ---------- ---------- ---------- ---------- ----------
NA_G.sI     491  ---------- ---------- ---------- ---------- ----------
C_G.sI(V)   491  ---------- ---------- ---------- ---------- ----------
C_G.sIIa    466  ---------- ---------- ---------- ---------- ----------
NA_G.sIIa   468  ---------- ---------- ---------- ---------- ----------
CS_G.sIIa   458  ---------- ---------- ---------- ---------- ----------
CA_G.sIIa   467  ---------- ---------- ---------- ---------- ----------
CA_G.sIIb   460  ---------- ---------- ---------- ---------- ----------
NA_G.sIIb   460  ---------- ---------- ---------- ---------- ----------
NA_G.cal    459  ---------- ---------- ---------- ---------- ----------
C_O.uni     407  ---------- ---------- ---------- ---------- ----------
CS_O.uni    407  ---------- ---------- ---------- ---------- ----------
CA_O.uni    408  ---------- ---------- ---------- ---------- ----------
M_O.uni(V)  408  ---------- ---------- ---------- ---------- ----------
C_G.sac     437  ---------- ---------- ---------- ---------- ----------
CS_G.sac    437  ---------- ---------- ---------- ---------- ----------
C_G.rub,P   439  ---------- ---------- ---------- ---------- ----------
NA_G.rub,P  439  ---------- ---------- ---------- ---------- ----------
NA_G.rub,WI 432  ---------- ---------- ---------- ---------- ----------
CS_G.rub,WI 432  ---------- ---------- ---------- ---------- ----------
C_G.rub(P)  436  ---------- ---------- ---------- ---------- ----------
NA_G.rub,WII 425 ---------- ---------- ---------- ---------- ----------
C_G.rub,WII 425  ---------- ---------- ---------- ---------- ----------
C_G.con(P)  441  ---------- ---------- ---------- ---------- ----------
CS_G.con    440  ---------- ---------- ---------- ---------- ----------
NA_N.pac(D) 472  ---------- ---------- ---------- ---------- ----------
AA_N.pac(D) 472  ---------- ---------- ---------- ---------- ----------
C_N.dut     424  ---------- ---------- ---------- ---------- ----------
CS_G.bul,Ia 405  ---------- ---------- ---------- ---------- ----------
CA_G.bul,IId 413 ---------- ---------- ---------- ---------- ----------
M_G.bul(V)  412  ---------- ---------- ---------- ---------- ----------
Na_G.bul,Ib 412  ---------- ---------- ---------- ---------- ----------
NA_G.falc   391  ---------- ---------- ---------- ---------- ----------
CS_G.falc   392  ---------- ---------- ---------- ---------- ----------
Na_G.bul,IIb 411 ---------- ---------- ---------- ---------- ----------
AA_G.bul,IIb 411 ---------- ---------- ---------- ---------- ----------
Na-G.bul,IIa 411 ---------- ---------- ---------- ---------- ----------
AA_G.bul,IIa 411 ---------- ---------- ---------- ---------- ----------
AA_G.bul,IIc 414 ---------- ---------- ---------- ---------- ----------
NA-T.qui,IIa 641 ---------- ---------- ---------- ---------- ----------
AA_T.qui,IIa 641 ---------- ---------- ---------- ---------- ----------
NA_T.qui,IIb 652 ---------- ---------- ---------- ---------- ----------
AA_T.qui,IIb 649 ---------- ---------- ---------- ---------- ----------
CS_T.qui,I  540  ---------- ---------- ---------- ---------- ----------
CS_G.glu    400  ---------- ---------- ---------- ---------- ----------
NA_G.uvu    413  ---------- ---------- ---------- ---------- ----------
A.becca_P   417  ---------- ---------- ---------- ---------- ----------
Troch_P     426  ---------- ---------- ---------- ---------- ----------
H.germ_P    361  ---------- ---------- ---------- ---------- ----------
Textu_P     410  ---------- ---------- ---------- ---------- ----------
Biger_P     422  ---------- ---------- ---------- ---------- ----------
Boliv_P     421  ---------- ---------- ---------- ---------- ----------
G.oper_P    409  ---------- ---------- ---------- ---------- ----------
A.trian_P   575  AATTTATTTT AATGATATTA AATTATATTA TATATATATG TGTATATGTG
As.rara_P   574  CATGAATTTA TTCGTGTATT TTATATATAT ATGTGATTAT AGTGAGTGTT
A.angu_P    352  ---------- ---------- ---------- ---------- ----------
E.acul_P    281  ---------- ---------- ---------- ---------- ----------
Quin_P      387  ---------- ---------- ---------- ---------- ----------
M.secan_P   366  ---------- ---------- ---------- ---------- ----------
P.pert_P    411  ---------- ---------- ---------- ---------- ----------
Allogrom_P  384  ---------- ---------- ---------- ---------- ----------
Align       198  ---------- ---------- ---------- ---------- ----------

C_G.sI      491  ---------- ---------- ---------- ---------- ----------
NA_G.sI     491  ---------- ---------- ---------- ---------- ----------
C_G.sI(V)   491  ---------- ---------- ---------- ---------- ----------
C_G.sIIa    466  ---------- ---------- ---------- ---------- ----------
```

```
NA_G.sIIa    468 ---------- ---------- ---------- ---------- ----------
CS_G.sIIa    458 ---------- ---------- ---------- ---------- ----------
CA_G.sIIa    467 ---------- ---------- ---------- ---------- ----------
CA_G.sIIb    460 ---------- ---------- ---------- ---------- ----------
NA_G.sIIb    460 ---------- ---------- ---------- ---------- ----------
 NA_G.cal    459 ---------- ---------- ---------- ---------- ----------
  C_O.uni    407 ---------- ---------- ---------- ---------- ----------
 CS_O.uni    407 ---------- ---------- ---------- ---------- ----------
 CA_O.uni    408 ---------- ---------- ---------- ---------- ----------
M_O.uni(V)   408 ---------- ---------- ---------- ---------- ----------
  C_G.sac    437 ---------- ---------- ---------- ---------- ----------
 CS_G.sac    437 ---------- ---------- ---------- ---------- ----------
 C_G.rub,P   439 ---------- ---------- ---------- ---------- ----------
NA_G.rub,P   439 ---------- ---------- ---------- ---------- ----------
NA_G.rub,WI  432 ---------- ---------- ---------- ---------- ----------
CS_G.rub,WI  432 ---------- ---------- ---------- ---------- ----------
 C_G.rub(P)  436 ---------- ---------- ---------- ---------- ----------
NA_G.rub,WII 425 ---------- ---------- ---------- ---------- ----------
C_G.rub,WII  425 ---------- ---------- ---------- ---------- ----------
 C_G.con(P)  441 ---------- ---------- ---------- ---------- ----------
  CS_G.con   440 ---------- ---------- ---------- ---------- ----------
NA_N.pac(D)  472 ---------- ---------- ---------- ---------- ----------
AA_N.pac(D)  472 ---------- ---------- ---------- ---------- ----------
  C_N.dut    424 ---------- ---------- ---------- ---------- ----------
CS_G.bul,Ia  405 ---------- ---------- ---------- ---------- ----------
CA_G.bul,IId 413 ---------- ---------- ---------- ---------- ----------
 M_G.bul(V)  412 ---------- ---------- ---------- ---------- ----------
Na_G.bul,Ib  412 ---------- ---------- ---------- ---------- ----------
 NA_G.falc   391 ---------- ---------- ---------- ---------- ----------
 CS_G.falc   392 ---------- ---------- ---------- ---------- ----------
Na_G.bul,IIb 411 ---------- ---------- ---------- ---------- ----------
AA_G.bul,IIb 411 ---------- ---------- ---------- ---------- ----------
Na-G.bul,IIa 411 ---------- ---------- ---------- ---------- ----------
AA_G.bul,IIa 411 ---------- ---------- ---------- ---------- ----------
AA_G.bul,IIc 414 ---------- ---------- ---------- ---------- ----------
NA-T.qui,IIa 641 ---------- ---------- ---------- ---------- ----------
AA_T.qui,IIa 641 ---------- ---------- ---------- ---------- ----------
NA_T.qui,IIb 652 ---------- ---------- ---------- ---------- ----------
AA_T.qui,IIb 649 ---------- ---------- ---------- ---------- ----------
 CS_T.qui,I  540 ---------- ---------- ---------- ---------- ----------
  CS_G.glu   400 ---------- ---------- ---------- ---------- ----------
  NA_G.uvu   413 ---------- ---------- ---------- ---------- ----------
 A.becca_P   417 ---------- ---------- ---------- ---------- ----------
   Troch_P   426 ---------- ---------- ---------- ---------- ----------
  H.germ_P   361 ---------- ---------- ---------- ---------- ----------
   Textu_P   410 ---------- ---------- ---------- ---------- ----------
   Biger_P   422 ---------- ---------- ---------- ---------- ----------
   Boliv_P   421 ---------- ---------- ---------- ---------- ----------
  G.oper_P   409 ---------- ---------- ---------- ---------- ----------
 A.trian_P   625 GTTTTTATAC GNGTTATATG ATTTTNTTTA TGTATTTAAT CATTTTGCAT
 As.rara_P   624 TTGGTTTTAT ACGTGTGGTG TATTTTATTT ATATATTTTA ATCATTTTAT
  A.angu_P   352 ---------- ---------- ---------- ---------- ----------
  E.acul_P   281 ---------- ---------- ---------- ---------- ----------
    Quin_P   387 ---------- ---------- ---------- ---------- ----------
 M.secan_P   366 ---------- ---------- ---------- ---------- ----------
  P.pert_P   411 ---------- ---------- ---------- ---------- ----------
Allogrom_P   384 ---------- ---------- ---------- ---------- ----------
     Align   198 ---------- ---------- ---------- ---------- ----------

   C_G.sI    491 ---------- ---------- ---------- ---------- -GGAA-AACT
  NA_G.sI    491 ---------- ---------- ---------- ---------- -GGAA-AACT
 C_G.sI(V)   491 ---------- ---------- ---------- ---------- -GGAA-AACT
 C_G.sIIa    466 ---------- ---------- ---------- ---------- -GGAA-AACT
NA_G.sIIa    468 ---------- ---------- ---------- ---------- -GGAA-AACT
CS_G.sIIa    458 ---------- ---------- ---------- ---------- -GGAA-AACT
CA_G.sIIa    467 ---------- ---------- ---------- ---------- -GGAA-AACT
CA_G.sIIb    460 ---------- ---------- ---------- ---------- -GGAA-AACT
NA_G.sIIb    460 ---------- ---------- ---------- ---------- -GGAA-AACT
 NA_G.cal    459 ---------- ---------- ---------- ---------- -GGAA-AACT
  C_O.uni    407 ---------- ---------- ---------- ---------- -GACA-AACT
 CS_O.uni    407 ---------- ---------- ---------- ---------- -GACA-AACT
 CA_O.uni    408 ---------- ---------- ---------- ---------- -GGCA-AACT
M_O.uni(V)   408 ---------- ---------- ---------- ---------- -GGCA-AACT
  C_G.sac    437 ---------- ---------- ---------- ---------- -GATA-AACT
 CS_G.sac    437 ---------- ---------- ---------- ---------- -GATA-AACT
 C_G.rub,P   439 ---------- ---------- ---------- ---------- -GATA-AACT
NA_G.rub,P   439 ---------- ---------- ---------- ---------- -GATA-AACT
```

```
NA_G.rub,WI   432  ---------- ---------- ---------- ---------- -GATA-AACT
CS_G.rub,WI   432  ---------- ---------- ---------- ---------- -GATA-AACT
 C_G.rub(P)   436  ---------- ---------- ---------- ---------- -GATA-AACT
NA_G.rub,WII  425  ---------- ---------- ---------- ---------- -GATA-AACT
C_G.rub,WII   425  ---------- ---------- ---------- ---------- -GATA-AACT
 C_G.con(P)   441  ---------- ---------- ---------- ---------- -GATA-AACT
    CS_G.con  440  ---------- ---------- ---------- ---------- -GATA-AACT
NA_N.pac(D)   472  ---------- ---------- ---------- ---------- -ACCA-AACT
AA_N.pac(D)   472  ---------- ---------- ---------- ---------- -ACCA-AACT
     C_N.dut  424  ---------- ---------- ---------- ---------- -TTAA-AATT
CS_G.bul,Ia   405  ---------- ---------- ---------- ---------- -TTAA-AACT
CA_G.bul,IId  413  ---------- ---------- ---------- ---------T TTTGA-AACT
 M_G.bul(V)   412  ---------- ---------- ---------- ---------- GTCGA-AACT
Na_G.bul,Ib   412  ---------- ---------- ---------- ---------- TTCGA-AACT
   NA_G.falc  391  ---------- ---------- ---------- ---------- -TTGT-AACT
   CS_G.falc  392  ---------- ---------- ---------- ---------- -TTGT-AACT
Na_G.bul,IIb  411  ---------- ---------- ---------- ---------T TTTGA-AACT
 AA_G.bul,IIb 411  ---------- ---------- ---------- ---------T TTTGA-AACT
Na-G.bul,IIa  411  ---------- ---------- ---------- ---------T TTTGA-AACT
 AA_G.bul,IIa 411  ---------- ---------- ---------- ---------T TTTGA-AACT
 AA_G.bul,IIc 414  ---------- ---------- ---------- ---------T TTTGA-AACT
NA-T.qui,IIa  641  ---------- ---------- ---------- ---------- -AATG-AACT
 AA_T.qui,IIa 641  ---------- ---------- ---------- ---------- -AATG-AACT
NA_T.qui,IIb  652  ---------- ---------- ---------- ---------- -AATG-AACT
 AA_T.qui,IIb 649  ---------- ---------- ---------- ---------- -AATG-AACT
 CS_T.qui,I   540  ---------- ---------- ---------- ---------- -AATG-AACT
   CS_G.glu   400  ---------- ---------- ---------- ---------- -CCAA-AACT
   NA_G.uvu   413  ---------- ---------- ---------- ---------- -AGTA-AACT
  A.becca_P   417  ---------- ---------- ---------- ---------- -ATTA-AACT
    Troch_P   426  ---------- ---------- ---------- ---------- -TTTA-AACT
   H.germ_P   361  ---------- ---------- ---------- ---------- -TATA-AACT
    Textu_P   410  ---------- ---------- ---------- ---------- -TTTA-AACT
    Biger_P   422  ---------- ---------- ---------- ---------- -TTTA-AACT
    Boliv_P   421  ---------- ---------- ---------- ---------- -CACA-AACT
   G.oper_P   409  ---------- ---------- ---------- ---------- -AATA-AACT
  A.trian_P   675  AAATTG---- ---------- ---------- ---------- -AGGNAAACG
  As.rara_P   674  ATAAATTG-- ---------- ---------- ---------- -AGGCTAACT
   A.angu_P   352  ---------- ---------- ---------- ---------- -CTTAAAATT
   E.acul_P   281  ---------- ---------- ---------- ---CGCGAGT ATATA-AACC
     Quin_P   387  ---------- ---------- ---------- ---------- -CTTAAAATT
  M.secan_P   366  ---------- ---------- ---------- ---------- -CTTAAAATT
   P.pert_P   411  ---------- ---------- ---------- ---------- -CATAAAATT
 Allogrom_P   384  ---------- ---------- ---------- ---------- -TGCT-AACT
      Align   198  ---------- ---------- ---------- ---------- ------mmmm

       C_G.sI  499  TGG----GCG ACCGCT---- ---------- --------GT AATACTTCTC
      NA_G.sI  499  TGG----GCG ACCGCT---- ---------- --------GT AATACTTCTC
    C_G.sI(V)  499  TGG----GCG ACCGCT---- ---------- --------GT AATACTTCTC
     C_G.sIIa  474  CGG----GCG ACCGCT---- ---------- --------GT AATATTTCTC
    NA_G.sIIa  476  CGG----GCG ACCGCT---- ---------- --------GT AATATTTCTC
    CS_G.sIIa  466  CGG----GCG ACCGCT---- ---------- --------GT AATATTTCTC
    CA_G.sIIa  475  CGG----GCG ACCGCT---- ---------- --------GT AATATTTCTC
    CA_G.sIIb  468  TGG----GCG ACCGCT---- ---------- --------GT AATACTTCTC
    NA_G.sIIb  468  TGG----GCG ACCGCT---- ---------- --------GT AATACTTCTC
     NA_G.cal  467  CGG----GCG ACCGCT---- ---------- --------GT AATACTTTCT
      C_O.uni  415  CGG----GGG ACAGCT---- ---------- --------T- CAATCAATTT
     CS_O.uni  415  CGG----GGG ACAGCT---- ---------- --------T- CAATCAATTT
     CA_O.uni  416  CAG----GGG ACAGCT---- ---------- --------T- CAACTATTCT
   M_O.uni(V)  416  CAG----GGG ACAGCT---- ---------- --------T- CAACTATTCT
      C_G.sac  445  TAA----GCG ACCGCT---- ---------- ------CCAA CATTGTGTTT
     CS_G.sac  445  TAA----GCG ACCGCT---- ---------- ------CCAA CATTGTGTTT
    C_G.rub,P  447  CGG----GGG ACTGCT---- ---------- -------GAC TATAACCATT
   NA_G.rub,P  447  CGG----GGG ACTGCT---- ---------- -------GAC TATAACCATT
  NA_G.rub,WI  440  CGG----GGG ACTGCT---- ---------- -------GAC TATAACCATT
  CS_G.rub,WI  440  CGG----GGG ACTGCT---- ---------- -------GAC TATAACCATT
   C_G.rub(P)  444  CGG----GGG ACTGCT---- ---------- -------GAC TATAACCATT
 NA_G.rub,WII  433  CGG----GGG ACTGCT---- ---------- -------GAC TATAAACTTT
 C_G.rub,WII   433  CGG----GGG ACTGCT---- ---------- -------GAC TATAAACTTT
   C_G.con(P)  449  CGG----GGG ACTGCT---- ---------- -------GAC TATAACCATT
     CS_G.con  448  CGG----GGG ACTGCT---- ---------- -------GAC TATAACCATT
NA_N.pac(D)    480  AGG----CGT ACCGCT---- ---------- ---------- ---GTATCAT
AA_N.pac(D)    480  AGG----CGT ACCGCT---- ---------- ---------- ---GTATCAT
     C_N.dut   432  AGA----CGG ACCGCT---- ---------- ---------- ---GTACTTT
CS_G.bul,Ia    413  CGA----GAA ACATCT---- ---------- ---------- ---GTGACTT
CA_G.bul,IId   423  CGG----GGA ACATCT---- ---------- ---------- ---GTGACTT
 M_G.bul(V)    421  CGG----GAA ACATCT---- ---------- ---------- ---GTGACTT
Na_G.bul,Ib    421  CGG----GAA ACATCT---- ---------- ---------- ---GTGACTT
```

```
   NA_G.falc   399  CGG----GGG ACCGCT---- ---------- ---------- ---GTGATTT
   CS_G.falc   400  CGG----GGG ACCGCT---- ---------- ---------- ---GTGATTT
Na_G.bul,IIb   421  CGG----GGA ACATCT---- ---------- ---------- ---GTGACTT
  AA_G.bul,IIb 421  CGG----GGA ACATCT---- ---------- ---------- ---GTGACTT
Na-G.bul,IIa   421  CGG----GGA ACATCT---- ---------- ---------- ---GTGACTT
  AA_G.bul,IIa 421  CGG----GGA ACATCT---- ---------- ---------- ---GTGACTT
  AA_G.bul,IIc 424  CGG----GGA ACATCT---- ---------- ---------- ---GTGACTT
NA-T.qui,IIa   649  CAG----GCG ACTGCT---- ---------- ---------- ----ATACTT
  AA_T.qui,IIa 649  CAG----GCG ACTGCT---- ---------- ---------- ----ATACTT
NA_T.qui,IIb   660  CAG----GCG ACTGCT---- ---------- ---------- ----ATACTT
  AA_T.qui,IIb 657  CAG----GCG ACTGCT---- ---------- ---------- ----ATACTT
  CS_T.qui,I   548  TAG----GCG ACTGCT---- ---------- ---------- ----ATACTT
    CS_G.glu   408  AGA----GGG ACCGCT---- ---------- ---------- ---GTCAACT
    NA_G.uvu   421  AGA----GGG ACCGCT---- ---------- ---------- ---GTCAACT
   A.becca_P   425  ATA----GAG ACCGCT---- ---------- ---------- ---GTTTCTT
     Troch_P   434  AGA----GGG ACCGCT---- ---------- ---------- ---GTAATCT
    H.germ_P   369  AGA----GGG ACCGCT---- ---------- ---------- ---GTTACTT
     Textu_P   418  AGA----GGG ACCGCT---- ---------- ---------- ---GTAATCT
     Biger_P   430  AGA----GGG ACCGCT---- ---------- ---------- ---GTAATCT
     Boliv_P   429  AGA----GGG ACCGCT---- ---------- ---------- ---GTTACTT
    G.oper_P   417  AGA----GGG ACCGCT---- ---------- ---------- ---GTTACTT
   A.trian_P   690  AGA----GGG ACCGCT---- ---------- ---------- -AGGCTAGCT
   As.rara_P   691  AGA----GGG ACCGCT---- ---------- ---------- -AGGCTAGCT
    A.angu_P   361  AAA----GGG ACCGCT---- ---------- ---------- GTCATTATTC
    E.acul_P   297  AAAATGTGAG ACCGCT---- ---------- ---------- ---GTTTCTT
      Quin_P   396  AAA----GGA ACCGCT---- ---------- ---------- GTCTAATATT
   M.secan_P   375  AAG----GGA ACCGCT---- ---------- ---------- GTCTAATATT
    P.pert_P   420  AAA----GGG ACCGCT---- ---------- ---------- GTCTAAATTT
  Allogrom_P   392  AGA----TGG ACCGCT---- ---------- ---------- ---G-GATCT
       Align   202  mmm----mmm mmmmmm---- ---------- ---------- ----------

      C_G.sI   523  TCT------T -AA-ACCA-- ---------- ---------- -GAGGAAGAT
     NA_G.sI   523  TCT------T -AA-ACCA-- ---------- ---------- -GAGGAAGAT
   C_G.sI(V)   523  TCT------T -AA-ACCA-- ---------- ---------- -GAGGAAGAT
    C_G.sIIa   498  T-T------T -AA-AACA-- ---------- ---------- -GAGGAAGAT
   NA_G.sIIa   500  T-T------T -AA-AACA-- ---------- ---------- -GAGGAAGAT
   CS_G.sIIa   490  T-T------T -AA-AACA-- ---------- ---------- -GAGGAAGAT
   CA_G.sIIa   499  T-T------T -AA-AACA-- ---------- ---------- -GAGGAAGAT
   CA_G.sIIb   492  T-T------T -AA-AACA-- ---------- ---------- -GAGGCAGGT
   NA_G.sIIb   492  T-T------T -AA-AACA-- ---------- ---------- -GAGGCAGGT
    NA_G.cal   491  TTT------T -AA-AACA-- ---------- ---------- -GAGGAAGAT
     C_O.uni   438  TCT------C -AA-ACGA-- ---------- ---------- -GAGGAAGGT
    CS_O.uni   438  TCT------C -AA-ACGA-- ---------- ---------- -GAGGAAGGT
    CA_O.uni   439  TCTT-----C -AA-ACTA-- ---------- ---------- -GAGGAAGGT
  M_O.uni(V)   439  TCTT-----C -AA-ACTA-- ---------- ---------- -GAGGAAGGT
     C_G.sac   471  TTT------T -AA-AACT-- ---------- ---------- -GATGAAGGT
    CS_G.sac   471  TTT------T -AA-AACT-- ---------- ---------- -GATGAAGGT
   C_G.rub,P   472  TCT------C -AA-AACA-- ---------- ---------- -GAGGAAGGT
  NA_G.rub,P   472  TCT------C -AA-AACA-- ---------- ---------- -GAGGAAGGT
 NA_G.rub,WI   465  TCT------C -AA-AACA-- ---------- ---------- -GAGGAAGGT
 CS_G.rub,WI   465  TCT------C -AA-AACA-- ---------- ---------- -GAGGAAGGT
  C_G.rub(P)   469  TCT------C -AA-AACA-- ---------- ---------- -GAGGAAGGT
NA_G.rub,WII   458  TCT------C -AA-ACCA-- ---------- ---------- -GAGGAAGGT
 C_G.rub,WII   458  TCT------C -AA-ACCA-- ---------- ---------- -GAGGAAGGT
  C_G.con(P)   474  TCT------C -AA-AACA-- ---------- ---------- -GAGGAAGGT
    CS_G.con   473  TCT------C -AA-AACA-- ---------- ---------- -AAGGAAGGT
  NA_N.pac(D)  499  TTCTT----T -AA-ACCA-- ---------- ---------- -GAGGAAGGT
  AA_N.pac(D)  499  TTCTT----T -AA-ACCA-- ---------- ---------- -GAGGAAGGT
     C_N.dut   451  TCT------T -AA-ACCA-- ---------- ---------- -CAGGAAGGT
  CS_G.bul,Ia  432  TCTTTCT--T -TA-CGCA-- ---------- ---------- -GAGGAAGGT
 CA_G.bul,IId  442  TCTTTCT--T -AA-CGCA-- ---------- ---------- -GAGGAAGGT
   M_G.bul(V)  440  TCTTTCT--T -TA-CGCA-- ---------- ---------- -GAGGAAGGT
  Na_G.bul,Ib  440  TCTTTCT--T -TA-CGCA-- ---------- ---------- -GAGGAAGGT
    NA_G.falc  418  TTCTT----C -AA-ACCG-- ---------- ---------- -GAGGAAGGT
    CS_G.falc  419  TTCTT----C -AA-ACCG-- ---------- ---------- -GAGGAAGGT
 Na_G.bul,IIb  440  TCTTTCT--T -AA-CGCA-- ---------- ---------- -GAGGAAGGT
  AA_G.bul,IIb 440  TCTTTCT--T -AA-CGCA-- ---------- ---------- -GAGGAAGGT
Na-G.bul,IIa   440  TCTTTCT--T -AA-CGCA-- ---------- ---------- -GAGGAAGGT
  AA_G.bul,IIa 440  TCTTTCT--T -AA-CGCA-- ---------- ---------- -GAGGAAGGT
  AA_G.bul,IIc 443  TCTTTCT--T -AA-CGCA-- ---------- ---------- -GAGGAAGGT
NA-T.qui,IIa   667  ---------T -AA-GATG-- ---------- ---------- -GTGGAAGGT
  AA_T.qui,IIa 667  ---------T -AA-GATG-- ---------- ---------- -GTGGAAGGT
NA_T.qui,IIb   678  ---------T -AA-GATG-- ---------- ---------- -GTGGAAGGT
  AA_T.qui,IIb 675  ---------T -AA-GATG-- ---------- ---------- -GTGGAAGGT
  CS_T.qui,I   566  T--------T -AA-GATG-- ---------- ---------- -GTCGAAGGT
    CS_G.glu   427  TCT------T -AA-ACCA-- ---------- ---------- -GAGGAAGGT
    NA_G.uvu   440  TCT------T -AA-ACCA-- ---------- ---------- -GAGGAAGGT
```

```
A.becca_P   444  CTT------T -AA-ACCA-- ---------- ---------- -GAGGAAGGA
  Troch_P   453  TTT------T -AA-ACCA-- ---------- ---------- -GAGGAAGGT
 H.germ_P   388  TCTTT----T -AA-ACCA-- ---------- ---------- -GAGGAAGGT
  Textu_P   437  TTT------T -AA-ACCA-- ---------- ---------- -GAGGAAGGT
  Biger_P   449  TTT------T -AA-ACCA-- ---------- ---------- -GAGGAAGGT
  Boliv_P   448  TCT------T -AA-ACCG-- ---------- ---------- -GAGGAAGGT
 G.oper_P   436  TCT------T -AA-ACCA-- ---------- ---------- -GAGGAAGGT
A.trian_P   711  TTT------T -AA-AACA-- ---------- ---------- -GAGGAAGGT
As.rara_P   712  TTT------T -AA-AACA-- ---------- ---------- -GAGGAAGGT
 A.angu_P   383  ATATGTGT-T -AA-AATA-- ---------- ---------- -GAGTAAGAT
 E.acul_P   320  TCTTT----T -AA-ACCA-- ---------- ---------- -GAGGAAGGT
   Quin_P   418  TAGTTT---T -AA-AACA-- ---------- ---------- -GAGGAAGAT
M.secan_P   397  TAGTGT---T -AA-AACA-- ---------- ---------- -GAGGAAGAT
 P.pert_P   442  TAGTGT---T -AA-AATA-- ---------- ---------- -GAGTAAGAT
Allogrom_P  410  TTT------C -TA-AACA-- ---------- ---------- -GAGGAAGAT
    Align   214  ---------m -mm-mmmm-- ---------- ---------- -mmmmmmmmmm

    C_G.sI  542  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC CGGG-CCGCA
   NA_G.sI  542  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC CGGG-CCGCA
 C_G.sI(V)  542  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC CGGG-CCGCA
  C_G.sIIa  516  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC CGGG-CCGCA
 NA_G.sIIa  518  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC CGGG-CCGCA
 CS_G.sIIa  508  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC CGGG-CCGCA
 CA_G.sIIa  517  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC CGGG-CCGCA
 CA_G.sIIb  510  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC CGGG-CCGCA
 NA_G.sIIb  510  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC CGGG-CCGCA
  NA_G.cal  510  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC CGGG-CCGCA
    C_O.uni 457  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC TGGG-CCGCA
   CS_O.uni 457  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC TGGG-CCGCA
   CA_O.uni 459  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTTC TAGG-CTGCA
 M_O.uni(V) 459  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTTC TAGG-CTGCA
    C_G.sac 490  TGCGGCAATA ACAGGTCTG- TGATGCCC-T TCG-ATGTCC TGGG-CTGCA
   CS_G.sac 490  TGCGGCAATA ACAGGTCTG- TGATGCCC-T TCG-ATGTCC TGGG-CTGCA
  C_G.rub,P 491  TGCGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTCC CGGG-CTGCA
 NA_G.rub,P 491  TGCGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTCC CGGG-CTGCA
NA_G.rub,WI 484  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC CGGG-CTGCA
CS_G.rub,WI 484  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC CGGG-CTGCA
  C_G.rub(P) 488  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC CGGG-CTGCA
NA_G.rub,WII 477  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTTC CGGG-CTGCA
 C_G.rub,WII 477  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTTC CGGG-CTGCA
  C_G.con(P) 493  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC CGGG-CTGCA
   CS_G.con  492  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC CGGG-CTGCA
 NA_N.pac(D) 520  TGCGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTTC TGGG-CTGCA
 AA_N.pac(D) 520  TGCGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTTC TGGG-CTGCA
    C_N.dut  470  TGCGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTTC CGGG-CTGCA
 CS_G.bul,Ia 455  TATGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC AGAG-CTGCA
CA_G.bul,IId 465  TATGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTCC AGGG-CTGCA
  M_G.bul(V) 463  TATGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC AAAG-CTGCA
 Na_G.bul,Ib 463  TATGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC AAAG-CTGCA
   NA_G.falc 439  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC TGGG-CTGCA
   CS_G.falc 440  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTCC TGGG-CTGCA
Na_G.bul,IIb 463  TATGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTCC AGGG-CTGCA
 AA_G.bul,IIb 463  TATGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTCC AGGG-CTGCA
Na-G.bul,IIa 463  TATGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTCC AGGG-CTGCA
 AA_G.bul,IIa 463  TATGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTCC AGGG-CTGCA
 AA_G.bul,IIc 466  TATGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTCT AGGG-CTGCA
NA-T.qui,IIa 683  TGTGGCAATG ACAGGTCTG- TGATGCCCTT TAG-ATGTTC AAGG-CTGCA
 AA_T.qui,IIa 683  TGTGGCAATG ACAGGTCTG- TGATGCCCTT TAG-ATGTTC AAGG-CTGCA
NA_T.qui,IIb 694  TGTGGCAATG ACAGGTCTG- TGATGCCCTT TAG-ATGTTC AAGG-CTGCA
 AA_T.qui,IIb 691  TGTGGCAATG ACAGGTCTG- TGATGCCCTT TAG-ATGTTC AAGG-CTGCA
 CS_T.qui,I  583  TGTGGCAATG ACAGGTCTG- TGATGCCC-T TAG-ATGTTC AGGG-CTGCA
   CS_G.glu  446  TGCGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTTC CGGG-CTGCA
    NA_G.uvu 459  TGCGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTTC CGGG-CTGCA
  A.becca_P  463  TACGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTTC CGGG-CTGCA
    Troch_P  472  TGCGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTTC CGGG-CTGCA
   H.germ_P  409  TGCGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTTC CGGG-CTGCA
    Textu_P  456  TGCGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTTC CGGG-CTGCA
    Biger_P  468  TGCGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTTC CGGG-CTGTA
    Boliv_P  467  TGCGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTTC CGGG-CTGCA
   G.oper_P  455  TGCGGCAATA ACAGGTCTG- TGATGCCC-T TAG-ATGTTC CGGG-CTGCA
  A.trian_P  730  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTTC CGGG-CTGCA
  As.rara_P  731  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTTC CGGG-CTG-A
   A.angu_P  407  TACGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTTC TGGG-CTGCA
   E.acul_P  341  TATGGCAATA ACAGGTCTG- TGATGCCC-T TCG-ATGTTC CGGG-CTGCA
     Quin_P  440  TATGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTTC TGGG-CTGCA
  M.secan_P  419  TATGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTTC TGGG-CTGCA
   P.pert_P  464  TACGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTTC TGGG-CTGCA
```

```
Allogrom_P   429  TGCGGCAATA ACAGGTCTG- TGATGCCC-T CAG-ATGTTC TGGG-CTGCA
     Align   230  mmmmmmmmmmm mmmmmmmmmm- mmmmmmmmm-m mmm-mmmmmm mmmm-mmmmm

     C_G.sI  588  CACGTGCTAC ATTGA-TT-A GCG-CAGTGC GCATACTTAC A-TAT-----
    NA_G.sI  588  CACGTGCTAC ATTGA-TT-A GCG-CAGTGC GCATACTTAC A-TAT-----
  C_G.sI(V)  588  CACGTGCTAC ATTGA-TT-A GCG-CAGTGC GCATACTTAC A-TAT-----
   C_G.sIIa  562  CACGTGCTAC ATTGA-TT-A GCG-CAGTGC GCATACTTTC AATAT-----
  NA_G.sIIa  564  CACGTGCTAC ATTGA-TT-A GCG-CAGTGC GCATACTTTC AATAT-----
  CS_G.sIIa  554  CACGTGCTAC ATTGA-TT-A GCG-CAGTGC GCATACTTTC AATAT-----
  CA_G.sIIa  563  CACGTGCTAC ATTGA-TT-A GCG-CAGTGC GCATACTTTC AATAT-----
  CA_G.sIIb  556  CACGTGCTAC ATTGA-TT-A GCG-CAGTGC GCATACTTTC AATAT-----
  NA_G.sIIb  556  CACGTGCTAC ATTGA-TT-A GCG-CAGTGC GCATACTTTC AATAT-----
   NA_G.cal  556  CACGTGCTAC ATTGA-TT-A GCG-CAGTGC GCATACTTTC AATAT-----
     C_O.uni 503  CACGTGCTAC ATTGA-TC-A GCG-CAGTGC GTCTTTTTTC C----------
    CS_O.uni 503  CACGTGCTAC ATTGA-TC-A GCG-CAGTGC GTCTTTTTTC C----------
    CA_O.uni 505  CACGTGCTAC ATTGA-TC-A GCA-CAATGC GATACATTTC G----------
  M_O.uni(V) 505  CACGTGCTAC ATTGA-TC-A GCA-CAATGC GATACATTTC G----------
     C_G.sac 536  CACGTGCTAC AATGG-CT-A GCG-CAGTGA GCATATTTCA ----------
    CS_G.sac 536  CACGTGCTAC AATGG-CT-A GCG-CAGTGA GCATATTTCA ----------
   C_G.rub,P 537  CACGTGCTAC AATGA-GT-G GCG-CAATGT GTGCAATTTT GT---------
  NA_G.rub,P 537  CACGTGCTAC AATGA-GT-G GCG-CAATGT GTGCAATTTT GT---------
 NA_G.rub,WI 530  CACGTGCTAC AATGA-GT-G GCG-CAATGT GCTAATGTTT ----------
 CS_G.rub,WI 530  CACGTGCTAC AATGA-GT-G GCG-CAATGT GCTAATGTTT ----------
   C_G.rub(P) 534  CACGTGCTAC AATGA-GTTG GCG-CAATGT GCTAATGTTT ----------
NA_G.rub,WII 523  CACGTGCTAC ATTTA-GT-A GCG-CAGTGT GCACATTAAT ----------
 C_G.rub,WII 523  CACGTGCTAC ATTTA-GT-A GCG-CAGTGT GCACATTAAT ----------
   C_G.con(P) 539  CGCGTGCTAC AATGA-GT-A GCG-CAGTGT GCACATTTT- ----------
    CS_G.con  538  CACGTGCTAC AATGA-GT-A GCG-CAGTGT GCACATTTT- ----------
  NA_N.pac(D) 566  CACGTGCTAC AATGA-TC-A GCA-CAGTAA GCATCTCAAT TTTTA-----
  AA_N.pac(D) 566  CACGTGCTAC AATGA-TC-A GCA-CAGTAA GCATCTCAAT TTTTA-----
     C_N.dut  516  CACGTGCTAC AATGA-TC-A GTA-CAGTGA GCATCTCAAT ATTAT-----
 CS_G.bul,Ia  501  CACGTACTAC AGTGA-TC-G ACT-CACTAA GTGTCGTGTT TTCTCCA---
CA_G.bul,IId  511  CACGTACTAC ATTGA-TC-A ACT-CAGTAG GCGTCGTGTT AGTTCTCC--
   M_G.bul(V) 509  CACGTACTAC AGTGA-TC-G TAA-CACTTG -CGTTGTTGT TAGCTCCA--
 Na_G.bul,Ib  509  CACGTACTAC AGTGA-TC-G TAA-CACTTG GCGTTGTTGT TAGCTCCA--
    NA_G.falc 485  CACGTGCTAC ACTGA-TC-C CGG-CAGTTG GCATGTGATT TTCTCCA---
    CS_G.falc 486  CACGTGCTAC ACTGA-TC-T CCA-CAGTTG GCATGTGATT TTCTCCA---
 Na_G.bul,IIb 509  CACGTACTAC ATTGA-TC-A ACT-CAGTAG GCGTCGTGTT AGTTCTCC--
   AA_G.bul,IIb 509  CACGTACTAC ATTGA-TC-A ACT-CAGTAG GCGTCGTGTT AGTTCTCC--
 Na-G.bul,IIa 509  CACGTACTAC ATTGA-TC-A ACT-CAGTAG GCGTCGTGTT TTCTCC----
   AA_G.bul,IIa 509  CACGTACTAC ATTGA-TC-A ACT-CAGTAG GCGTCGTGTT TTCTCC----
   AA_G.bul,IIc 512  CACGTACTAC ATTGA-TC-A ACT-CAGTAG GCGTCGTGCT TTCTCC----
 NA-T.qui,IIa 730  CACGTACTAC ATTGA-TC-T AGT-CAACGA GTATGTATGT AACATTG---
   AA_T.qui,IIa 730  CACGTACTAC ATTGA-TC-T AGT-CAACGA GTATGTATGT AACATTG---
 NA_T.qui,IIb 741  CACGTACTAC ATTGA-TC-T AGT-CAACGA GTATGTATGT AACATTG---
   AA_T.qui,IIb 738  CACGTACTAC ATTGA-TC-T AGT-CAACGA GTATGTATGT AACATTG---
   CS_T.qui,I 629  CACGTACTAC ATTGA-TC-T AGT-CAATAA GTATGTGTGT AAC--------
     CS_G.glu 492  CACGTGCTAC AATGA-TT-G TTG-CAGTGA GCATCTCAAT TTTTT-----
     NA_G.uvu 505  CACGTGCTAC AATGA-TC-G TAG-CAGTGA GCATCTTATT TGATTT----
    A.becca_P 509  CACGTGCTAC AATGA-TC-A TTG-CACTGT GCATCTAACC CAATG-----
      Troch_P 518  CACGTGCTAC AATGA-TT-A TTG-CAGTGA GCATCTCATT TTATT-----
     H.germ_P 455  CACGTGCTAC AATGA-TC-A TTG-CAGTGC GCATCTCATT TGTTATA---
      Textu_P 502  CACGTGCTAC AATGA-TT-A TTG-CAGTGA GCATCTCATT TTTAACA---
      Biger_P 514  CACGTGCTAC AATGA-TT-A TTG-CAGTGA GCATCTCATT TATAACA---
      Boliv_P 513  CACGTGCTAC AATGA-TT-A TTG-CAGTGA GCATCTATTT TTTACTA---
     G.oper_P 501  CACGTGCTAC AATGA-TT-A TTG-CAGTGA GCATCTCATA TTTACTT---
    A.trian_P 776  CACGTGCTAC AATGA-TT-A TTG-CAGTGA GCATCTCAAC ATTGTGA---
    As.rara_P 776  CACGTGCTAC AATGA-TT-A TTG-CAGTGA GCATCTCAAC ATTGTGA---
     A.angu_P 453  CACGTGCTAC AATAA-TT-A CAT-TAATAA GTATATACAT TATATTT---
     E.acul_P 387  CACGTGCTAC AATGA-TT-A TAT-CATTAA GTACATATA- ----------
       Quin_P 486  CACGTGCTAC AATAA-TT-A CTA-TAATAA GTGTCTAATT ATTATAT---
    M.secan_P 465  CACGTGCTAC AATAA-TT-A CTA-TACTTA GTGTCTACTT TTTCATG---
     P.pert_P 510  CACGTGCTAC AATAA-TT-A CAT-TAATAA GTATCTATTA TAATACA---
   Allogrom_P 475  CACGTGCTAC AATGA-TT-A TTG-CAGTAA GCATCTATAT AGGAT-----
        Align 276  mmmmmmmmmmm mmmmmm-mm-m mmm-mmmmmm mm-------- ----------

     C_G.sI  629  ---------- ------AG-A TCATT----G -ATT--G-G- TTATTT---A
    NA_G.sI  629  ---------- ------AG-A TCATT----G -ATT--G-G- TTATTT---A
  C_G.sI(V)  629  ---------- ------AG-A TCATT----G -ATT--G-G- TTATTT---A
   C_G.sIIa  604  ---------- ------GA-T ACATT---AG -ATT--G-GA TTGTTCTG--
  NA_G.sIIa  606  ---------- ------GA-T ACATT---AG -ATT--G-GA TTGTTCTG--
  CS_G.sIIa  596  ---------- ------GA-T ACATT---AG -ATT--G-GA TAGCTT----
  CA_G.sIIa  605  ---------- ------GA-T ACATT---AG -ATT--G-GA TTGTTCTG--
  CA_G.sIIb  598  ---------- ------GA-T ACATT---GG -ATT--G-GT GGTGAATTGT
  NA_G.sIIb  598  ---------- ------GA-T ACATT---GG -ATT--G-GT GGTGAATTGT
   NA_G.cal  598  ---------- ------GA-T ACATT---GG -TTTT-G-GT AGTAAGCTAT
     C_O.uni 541  ---------- ------AATG --A-A-ACA- TC----G-G- TCTG------
```

```
   CS_O.uni   541  ---------- ------AATG --A-A-ACA- TC----G-G- TCTG------
   CA_O.uni   543  ---------- ------AAAG GAAACATCGG TTTTAAGGTA ACGTCGATTG
  M_O.uni(V)  543  ---------- ------AAAG GAAACATCGG TTTTAAGGTA ACGTCGATTG
    C_G.sac   573  ---------- ------ACCG --A-A-ACA- TC----G-G- T-TG------
   CS_G.sac   573  ---------- ------ACCG --A-A-ACA- TC----G-G- T-TG------
   C_G.rub,P  576  ---------- ------AAGG G-ATA-AGAG TGT---G-G- TCAGT----A
  NA_G.rub,P  576  ---------- ------AAGG G-ATA-AGAG TGT---G-G- TCAGT----A
 NA_G.rub,WI  567  ---------- ------AAGG G-ATA-AGAG TGT---G-G- TCAGT----A
 CS_G.rub,WI  567  ---------- ------AAGG G-ATA-AGAG TGT---G-G- TCAGT----A
   C_G.rub(P) 572  ---------- -----TAAGG G-ATA-AGCG TGC---G-G- CCAGT----A
 NA_G.rub,WII 560  ---------- -----AGAGT G-ATA-GGTT CCGTGGTTTG TACATTTTAA
  C_G.rub,WII 560  ---------- -----AGAGT G-ATA-GGTT CCGTGGTTTG TACATTTTAA
   C_G.con(P) 575  ---------- ------GAAG AGTGAATAAG GTGCTGGG-- TGTACATTT-
    CS_G.con  574  ---------- ------GAAG AGTGA-TA-G GTGCTGGGGT TGTACATTTT
  NA_N.pac(D) 608  ---------- ------CAAC ACCGTCAACA CACGTAGTGA GCT-GCTTGA
  AA_N.pac(D) 608  ---------- ------CAAC ACCGTCAACA CACGTAGTGA GCT-GCTTGA
    C_N.dut   558  ---------- ------ACAC CGTATTAAGC GCTTAGTTGC GATTATTGGC
  CS_G.bul,Ia 545  ---------- ------AATA ACGTATACAG TGGACTTGGT GTCGGGTGC-
 CA_G.bul,IId 556  ---------- ------AATA ACGTATCTAG TGGACTTGGT GTCGGGTGTG
   M_G.bul(V) 553  ---------- -------TTA ACGTATTGAG TGGACTTGGT GTCGGGTC--
  Na_G.bul,Ib 554  ---------- -------TTA ACGTATTGAG TGGACTTGGT GTCGGGTGC-
    NA_G.falc 529  ---------- -------ATG ACGAACCTAG TGGACTTGGC GTCTGTGCGC
    CS_G.falc 530  ---------- -------ATG ACGAACCTAG TGGACTTGGC GTCTGTGCGC
 Na_G.bul,IIb 554  ---------- ------AATA ACGTATCTAG TGGACTTGGT GTCGGGTATG
  AA_G.bul,IIb 554  ---------- ------AATA ACGTATCTAG TGGACTTGGT GTCGGGTATG
 Na-G.bul,IIa 552  ---------- ------AATA ACGTATCTAG TGGACTTGGT GTCGGGTGCG
  AA_G.bul,IIa 552  ---------- ------AATA ACGTATCTAG TGGACTTGGT GTCGGGTGCG
  AA_G.bul,IIc 555  ---------- ------AATA ACGTATCTAG TGGACTTGGT GTCGGGTGCG
 NA-T.qui,IIa 774  ---------- ------AATA ATGAATGTAT TGGTTAAGCT ATATTGTATT
  AA_T.qui,IIa 774  ---------- ------AATA ATGAATGTAT TGGTTAAGCT ATATTGTATT
 NA_T.qui,IIb 785  ---------- ------ATTT TGCATCTATT GGTTAAGCTA TAATGTATTT
  AA_T.qui,IIb 782  ---------- ------ATTA TGCATCTATT GGTTAAGCTA TAATGTATTT
  CS_T.qui,I  669  ---------- ---AATAATT TTGAATGTAT TGGTTAAGCT TTTCTTATAT
    CS_G.glu  534  ---------- ------ACCT AACACCGCAC ACGTGAGTGC ATACTTGTAT
    NA_G.uvu  548  ---------- ------ACTT AACACCGCAT ACGTGAGTTC CAACTAGCTT
   A.becca_P  551  ---------- ------TGCG TGGACGCCGC ATGTTTGTGC CTTCGGGTAC
     Troch_P  560  ---------- ------ACAC ACAGCCT-GC GCGTGTCCAA TTATATACAA
    H.germ_P  499  ---------- ------CACT GCTTGTGCGT ATGTGCACCA TATATTTATA
     Textu_P  546  ---------- ---------- CACCGCTTGC GCGTGTCCGT AATATTTGTG
     Biger_P  558  ---------- ---------- CACCGCATGC GCGTGTCCAT AATATTTGCG
     Boliv_P  557  ---------- ------CACC GCATGCGCGA GTCTATTTGT CGCTTCACAG
    G.oper_P  545  ---------- ------ACAT CACTTGCATG CAGCACCTTA ACTTTTGTTA
   A.trian_P  820  ---------- ------TTTT AGTTTATTAT GATTATATAT TTATATATTA
   As.rara_P  820  ---------- ------TTTT AGATTATTAA AATAATTATA TTATGAAAAG
    A.angu_P  497  ---------- ------ATTA TATAAATTAA ---------- ----------
    E.acul_P  423  ---------- ---------- ---------- ---------- ----------
     Quin_P   530  ---------- ------AAAA ATTATTCATA TTTTATTATG TTTAATTTTT
   M.secan_P  509  ---------- ------TAAT ATATTTAACT ATATTTTTTA GTTTTATATA
    P.pert_P  554  ---------- ------TAAT ATGTTTACAT ATTATAATAA ----------
   Allogrom_P 517  ---------- ------TTTT ATAATCCGCA GGAATTTAAA TAAATATATA
       Align  305  ---------- ---------- ---------- ---------- ----------

     C_G.sI   650  TAACCG---- ---------- ---------- ---------- -------TCA
     NA_G.sI  650  TAACCG---- ---------- ---------- ---------- -------TCA
    C_G.sI(V) 650  TAACCG---- ---------- ---------- ---------- -------TCA
    C_G.sIIa  628  --------TT GAGCT-CCAT ---------- ---------- -------CTA
   NA_G.sIIa  630  --------TT GAGCT-CCAT ---------- ---------- -------CTA
   CS_G.sIIa  618  --------TT --GCTTCCAT ---------- ---------- -------CTA
   CA_G.sIIa  629  --------TT GAGCT-CCAT ---------- ---------- -------CTA
   CA_G.sIIb  624  TGGCCCCGTC TAATACT--- ---------- ---------- ----------
   NA_G.sIIb  624  TGGCCCCGTC TAATACT--- ---------- ---------- ----------
    NA_G.cal  625  TCATCTGATG GATATGCTAC TCTCCATACC AATACT---- ----------
     C_O.uni  558  -G-CCG--TT TTTAAC-AGT CCTGTACG-G TTT---TGCA ATGGCTACTG
    CS_O.uni  558  -G-CCG--TT TTTAAC-AGT CCTGTACG-G TTT---TGCA ATGGCTACTG
    CA_O.uni  577  CCTACCGATA GC-------- ---------- ---------- ----------
   M_O.uni(V) 577  CCTACCGATA GC-------- ---------- ---------- ----------
     C_G.sac  589  -G-CTG--TT CTGATTGACC CCTTCTGG-G TCTCTCTGTA ACA-CTACCG
    CS_G.sac  589  -G-CTG--TT CTGATTGACC CCTTCTGG-G TCTCTCTGTA ACA-CTACCG
    C_G.rub,P 599  TG-CCG--TC GGGGATTCCC CCTG-ACG-- TCTC-CTAGC ACA-CGATT-
   NA_G.rub,P 599  TG-CCG--TC GGGGATTCCC CCTG-ACG-- TCTC-CTAGC ACA-CGATT-
  NA_G.rub,WI 590  GA-ACG--TC GGG-ATTTCC CCTG-ACG-- TCTC-CCCCG CA--CGATT-
  CS_G.rub,WI 590  GA-ACG--TC GGG-ATTTCC CCTG-ACG-- TCTC-CCCCG CA--CGATT-
    C_G.rub(P) 596  GA-ACG--TC GGGAAATTCC CCTG--CG-- TCGCC-TGCG CA--CGATT-
  NA_G.rub,WII 593  GGGCTCGTCC TAGTAACCAA AGCTCCACGG GACATATCA- ----------
  C_G.rub,WII 593  GGGCTCGTCC TAGTAACCAA AGCTCCACGG GACATATCA- ----------
    C_G.con(P) 606  AAAGG-CTCG TCCTAGTAAC CAACCGCCCC CCAAGCATCT ATCA------
     CS_G.con  606  AAAGGGCTCG TCCTAGTAAC CAACCGCCCC C-AAGCATCT ATCA------
```

```
NA_N.pac(D)    641  TCTCTCATTC ATGCATTCTG TGCTTCGGTG CAGTGTGCAA TGTGAGTTGT
AA_N.pac(D)    641  TCTCTCATTC ATGCATTCTG TGCTTCGGTG CAGTGTGCAA TGTGAGTTGT
    C_N.dut    592  TCATTATTGG GTCTTTTAAT TGTATTTCT- --------AA TGCGCGCGGT
CS_G.bul,Ia    578  TGGCCTCTGG TCATGTGCTT TGATTAC--- ---------- ----------
CA_G.bul,IId   590  TGGCCTCTGG TCATGCGCTT TGATTAC--- ---------- ----------
  M_G.bul(V)   584  TGGCCTCTGG TCATGTGCTT TGATTACT-- ---------- ----------
 Na_G.bul,Ib   586  TGGCCTCTGG TCATGTGCTT TGATTACT-- ---------- ----------
   NA_G.falc   562  ATTCGTGTGC TAATGACTGC ---------- ---------- ----------
   CS_G.falc   563  ATTCGTGTGC TAATGACTGC ---------- ---------- ----------
Na_G.bul,IIb   588  TGGCCTC-GG TCATGTACTT TGATTAC--- ---------- ----------
  AA_G.bul,IIb 588  TGGCCTC-GG TCATGTACTT TGATTAC--- ---------- ----------
Na-G.bul,IIa   586  TGGCCTC-GG TCGTGTACTT TGATTAC--- ---------- ----------
  AA_G.bul,IIa 586  TGGCCTC-GG TCGTGTACTT TGATTAC--- ---------- ----------
  AA_G.bul,IIc 589  TGGCCTC-GG TCGTGTACTT TGATTAC--- ---------- ----------
NA-T.qui,IIa   808  TCGCTAACGT TAAT------ ---------- ---------- ----------
  AA_T.qui,IIa 808  TCGCTAACGT TAAT------ ---------- ---------- ----------
NA_T.qui,IIb   819  CGGTAACGTT AAT------- ---------- ---------- ----------
  AA_T.qui,IIb 816  CGGTAACGTT AAT------- ---------- ---------- ----------
  CS_T.qui,I   706  TTTGGTAAGA GGTTAAT--- ---------- ---------- ----------
    CS_G.glu   568  GTTACTTTAC GCAGCGGTAA A--------- ---------- ----------
    NA_G.uvu   582  AGCAATAAGT CAAGTAGATC TCTACGCAGC GGTAACA--- ----------
   A.becca_P   585  ACGCATTGCT GAGCGACCAC GCCGA----- ---------- ----------
     Troch_P   593  TTTATGTTGT ATTAATAATT G----TGT-G C--------- ---ATTGCGC
    H.germ_P   533  TGTGTGTGTA TGTATTGCAC GCAGTAA--- ---------- ----------
     Textu_P   576  CGTATTATGC GTATCTTATA TTGCGTGTTA C--------- ---ATTGCGC
     Biger_P   588  TGTTTATGCA TGTCTTATAT TATG-TGT-A C--------- ---ATTGTGC
     Boliv_P   591  CGCTCAAATA CGATCTCTGC GCGCGGTAAA ---------- ----------
    G.oper_P   579  TGGCGTCTTT ATGCGTGATA AA-------- ---------- ----------
   A.trian_P   854  TAATGTATTT ATTTATGTTA TTTTATATTA TATTAATTGT TTATTCTAAA
   As.rara_P   854  TGTATTTATT TATATTTTGA TTTTTATATT ATTTATAATT TAAAAATTTA
    A.angu_P   511  ---------- ---------- ---------- ---------- ----------
    E.acul_P   423  ---------- ---------- ---------- ---------- ----------
      Quin_P   564  TCAAA----- ---------- ---------- ---------- ----------
   M.secan_P   543  TTTTACATTA ---------- ---------- ---------- ----------
    P.pert_P   578  ---------- ---------- ---------- ---------- ----------
  Allogrom_P   551  ATTTTTTATA TATTTATCCT GTTTA----- ---------- ----------
       Align   305  ---------- ---------- ---------- ---------- ----------

       C_G.sI  659  ATATC----- ---------- ---------- ---------- ----------
      NA_G.sI  659  ATATC----- ---------- ---------- ---------- ----------
    C_G.sI(V)  659  ATATC----- ---------- ---------- ---------- ----------
     C_G.sIIa  642  ATACT----- ---------- ---------- ---------- ----------
    NA_G.sIIa  644  ATACT----- ---------- ---------- ---------- ----------
    CS_G.sIIa  631  ATACT----- ---------- ---------- ---------- ----------
    CA_G.sIIa  643  ATACT----- ---------- ---------- ---------- ----------
     CA_G.sIIb 641  ---------- ---------- ---------- ---------- ----------
    NA_G.sIIb  641  ---------- ---------- ---------- ---------- ----------
     NA_G.cal  661  ---------- ---------- ---------- ---------- ----------
      C_O.uni  599  ATAGT----- ---------- ---------- ---------- ----------
     CS_O.uni  599  ATAGT----- ---------- ---------- ---------- ----------
     CA_O.uni  589  ---------- ---------- ---------- ---------- ----------
   M_O.uni(V)  589  ---------- ---------- ---------- ---------- ----------
      C_G.sac  633  ATGAT----- ---------- ---------- ---------- ----------
     CS_G.sac  633  ATGAT----- ---------- ---------- ---------- ----------
    C_G.rub,P  640  ATCA------ ---------- ---------- ---------- ----------
   NA_G.rub,P  640  ATCA------ ---------- ---------- ---------- ----------
  NA_G.rub,WI  629  ATCA------ ---------- ---------- ---------- ----------
  CS_G.rub,WI  629  ATCA------ ---------- ---------- ---------- ----------
    C_G.rub(P) 635  ATCA------ ---------- ---------- ---------- ----------
 NA_G.rub,WII  632  ---------- ---------- ---------- ---------- ----------
  C_G.rub,WII  632  ---------- ---------- ---------- ---------- ----------
    C_G.con(P) 649  ---------- ---------- ---------- ---------- ----------
     CS_G.con  649  ---------- ---------- ---------- ---------- ----------
  NA_N.pac(D)  691  CTAATGCGCG CGGTTAT--- ---------- ---------- ----------
  AA_N.pac(D)  691  CTAATGCGCG CGGTTAT--- ---------- ---------- ----------
    C_N.dut    633  AAA------- ---------- ---------- ---------- ----------
 CS_G.bul,Ia   605  ---------- ---------- ---------- ---------- ----------
CA_G.bul,IId   617  ---------- ---------- ---------- ---------- ----------
  M_G.bul(V)   612  ---------- ---------- ---------- ---------- ----------
 Na_G.bul,Ib   614  ---------- ---------- ---------- ---------- ----------
   NA_G.falc   582  ---------- ---------- ---------- ---------- ----------
   CS_G.falc   583  ---------- ---------- ---------- ---------- ----------
Na_G.bul,IIb   614  ---------- ---------- ---------- ---------- ----------
  AA_G.bul,IIb 614  ---------- ---------- ---------- ---------- ----------
Na-G.bul,IIa   612  ---------- ---------- ---------- ---------- ----------
  AA_G.bul,IIa 612  ---------- ---------- ---------- ---------- ----------
  AA_G.bul,IIc 615  ---------- ---------- ---------- ---------- ----------
```

```
NA-T.qui,IIa    822  ---------- ---------- ---------- ---------- ----------
  AA_T.qui,IIa  822  ---------- ---------- ---------- ---------- ----------
NA_T.qui,IIb    832  ---------- ---------- ---------- ---------- ----------
  AA_T.qui,IIb  829  ---------- ---------- ---------- ---------- ----------
  CS_T.qui,I    723  ---------- ---------- ---------- ---------- ----------
    CS_G.glu    589  ---------- ---------- ---------- ---------- ----------
    NA_G.uvu    619  ---------- ---------- ---------- ---------- ----------
   A.becca_P    610  ---------- ---------- ---------- ---------- ----------
     Troch_P    626  GCTGTAAA-- ---------- ---------- ---------- ----------
    H.germ_P    560  ---------- ---------- ---------- ---------- ----------
     Textu_P    614  GCGGTAAA-- ---------- ---------- ---------- ----------
     Biger_P    624  GCGGTAAA-- ---------- ---------- ---------- ----------
     Boliv_P    621  ---------- ---------- ---------- ---------- ----------
    G.oper_P    601  ---------- ---------- ---------- ---------- ----------
   A.trian_P    904  AATTTAATCA CATA------ ---------- ---------- ----------
   As.rara_P    904  ATCACATA-- ---------- ---------- ---------- ----------
    A.angu_P    511  ---------- ---------- ---------- ---------- ----------
    E.acul_P    423  ---------- ---------- ---------- ---------- ----------
      Quin_P    569  ---------- ---------- ---------- ---------- ----------
   M.secan_P    553  ---------- ---------- ---------- ---------- ----------
    P.pert_P    578  ---------- ---------- ---------- ---------- ----------
  Allogrom_P    576  ---------- ---------- ---------- ---------- ----------
       Align    305  ---------- ---------- ---------- ---------- ----------

      C_G.sI    664  ---------- ---------- ---------- ---GCCTTGT CT--GA-AA-
     NA_G.sI    664  ---------- ---------- ---------- ---GCCTTGT CT--GA-AA-
    C_G.sI(V)   664  ---------- ---------- ---------- ---GCCTTGT CT--GA-AA-
     C_G.sIIa   647  ---------- ---------- ---------- ---ATCCGGC TT--GA-GA-
    NA_G.sIIa   649  ---------- ---------- ---------- ---ATCCGGC TT--GA-GA-
    CS_G.sIIa   636  ---------- ---------- ---------- ---ATCCGGC TT--GA-GA-
    CA_G.sIIa   648  ---------- ---------- ---------- ---ATCCGGC TT--GA-GA-
    CA_G.sIIb   641  ---------- ---------- ---------- ---GTCCTGT CT--GA-GA-
    NA_G.sIIb   641  ---------- ---------- ---------- ---GTCCTGT CT--GA-GA-
     NA_G.cal   661  ---------- ---------- ---------- ---ATCTAGT CT--GA-AA-
      C_O.uni   604  ---------- ---------- ---------- ---GTCTGGT GTT-CA-AC-
     CS_O.uni   604  ---------- ---------- ---------- ---GTCTGGT GTT-CA-AC-
     CA_O.uni   589  ---------- ---------- ---------- ---ACCTGGT CCC-CA-GT-
    M_O.uni(V)  589  ---------- ---------- ---------- ---ACCTGGT CCC-CA-GT-
      C_G.sac   638  ---------- ---------- ---------- ---CCCTCCT CT--GA-AA-
     CS_G.sac   638  ---------- ---------- ---------- ---CCCTCCT CT--GA-AA-
    C_G.rub,P   644  ---------- ---------- ---------- ---GCCTTTC CT--TA-AC-
    NA_G.rub,P  644  ---------- ---------- ---------- ---GCCTTTC CT--TA-AC-
   NA_G.rub,WI  633  ---------- ---------- ---------- ---GCCTTCC CT--TA-AC-
   CS_G.rub,WI  633  ---------- ---------- ---------- ---GCCTTCC CT--TA-AC-
    C_G.rub(P)  639  ---------- ---------- ---------- ---GCCTTCC CT--TA-AC-
  NA_G.rub,WII  632  ---------- ---------- ---------- ---ACCTTCC CC--TT-AC-
   C_G.rub,WII  632  ---------- ---------- ---------- ---ACCTTCC CC--TT-AC-
    C_G.con(P)  649  ---------- ---------- ---------- ---ACCTTCC -T--TT-AC-
     CS_G.con   649  ---------- ---------- ---------- ---ACCTTCC CT--TT-AC-
   NA_N.pac(D)  708  ---------- ---------- ---------- ---GCCTGTA CC--GA-GA-
   AA_N.pac(D)  708  ---------- ---------- ---------- ---GCCTGTA CC--GA-GA-
      C_N.dut   636  ---------- ---------- ---------- ---GCCTGCT TC--GA-GA-
   CS_G.bul,Ia  605  ---------- ---------- ---------- ---TGTCACT TT--AA-AC-
  CA_G.bul,IId  617  ---------- ---------- ---------- ---TGTCACT TT--AA-AC-
    M_G.bul(V)  612  ---------- ---------- ---------- ---GAACACT TT--AA-AC-
   Na_G.bul,Ib  614  ---------- ---------- ---------- ---GAACACT TT--AA-AC-
     NA_G.falc  582  ---------- ---------- ---------- ---CGTCACT TT--AA-AC-
     CS_G.falc  583  ---------- ---------- ---------- ---CGTCACT TT--AA-AC-
   Na_G.bul,IIb 614  ---------- ---------- ---------- ---TGTCACT TT--AA-AC-
    AA_G.bul,IIb 614 ---------- ---------- ---------- ---TGTCACT TT--AA-AC-
  Na-G.bul,IIa  612  ---------- ---------- ---------- ---TGTCACT TT--AA-AC-
    AA_G.bul,IIa 612 ---------- ---------- ---------- ---TGTCACT TT--AA-AC-
    AA_G.bul,IIc 615 ---------- ---------- ---------- ---TGTCACT TT--AA-AC-
  NA-T.qui,IIa  822  ---------- ---------- ---------- ---TCAGAAC TT--CG-AG-
    AA_T.qui,IIa 822 ---------- ---------- ---------- ---TCAGAAC TT--CG-AG-
  NA_T.qui,IIb  832  ---------- ---------- ---------- ---TCAGAAC TT--CG-AG-
    AA_T.qui,IIb 829 ---------- ---------- ---------- ---TCAGAAC TT--CG-AG-
    CS_T.qui,I  723  ---------- ---------- ---------- ---ACAGAAC TT--CG-AG-
     CS_G.glu   589  ---------- ---------- ---------- ---GCCTGCT TC--GA-GA-
     NA_G.uvu   619  ---------- ---------- ---------- ---GCCTGCT TC--GA-GA-
    A.becca_P   610  ---------- ---------- ---------- ---ACCTACT TC--GA-AA-
      Troch_P   634  ---------- ---------- ---------- ---GCCTGCT TC--GA-AA-
     H.germ_P   560  ---------- ---------- ---------- ---AGCCTAC TT--CG-AA-
      Textu_P   622  ---------- ---------- ---------- ---GCCTACT TC--GA-AA-
      Biger_P   632  ---------- ---------- ---------- ---GCCTACT TC--GA-AA-
      Boliv_P   621  ---------- ---------- ---------- ---GCCTACT TC--GA-AA-
     G.oper_P   601  ---------- ---------- ---------- ---GCCTGCT TC--GA-AA-
```

```
A.trian_P   918  ---------- ---------- ---------- ---GCCTACT TC--GG-CA-
As.rara_P   912  ---------- ---------- ---------- ---GCCTACT TC--GG-CA-
A.angu_P    511  ---------- ---------- ---------- ---ACCTATT TC--GA-AA-
E.acul_P    423  ---------- ---------- ---------- ---TCCTACT CT--GA-GA-
Quin_P      569  ---------- ---------- ---------- ---ACCTATT TC--GA-AA-
M.secan_P   553  ---------- ---------- ---------- ---ACCTATT TC--GA-AA-
P.pert_P    578  ---------- ---------- ---------- ---ACCTATT TC--GA-AA-
Allogrom_P  576  ---------- ---------- ---------- ---ACCAACT TT--GA-AA-
Align       305  ---------- ---------- ---------- ---------- ----------

C_G.sI      677  ----GGAC-- ---------- ----TAGG-T AATCTATTGT AAGTGCTGGT
NA_G.sI     677  ----GGAC-- ---------- ----TAGG-T AATCTATTGT AAGTGCTGGT
C_G.sI(V)   677  ----GGAC-- ---------- ----TAGG-T AATCTATTGT AAGTGCTGGT
C_G.sIIa    660  ----AGGC-- ---------- ----TGGG-T AATCAATTGT AAGTGCTGGT
NA_G.sIIa   662  ----AGGC-- ---------- ----TGGG-T AATCAATTGT AAGTGCTGGT
CS_G.sIIa   649  ----AGGC-- ---------- ----TGGG-T AATCAATTGT AAGTGCTGGT
CA_G.sIIa   661  ----AGGC-- ---------- ----TGGG-T AATCAATTGT AAGTGCTGGT
CA_G.sIIb   654  ----AGGC-- ---------- ----TGGG-T AATCAATTGT AAGTGCTGGT
NA_G.sIIb   654  ----AGGC-- ---------- ----TGGG-T AATCAATTGT AAGTGCTGGT
NA_G.cal    674  ----AGAC-- ---------- ----TGGG-T AATCTATTGT AAGTGCTGGT
C_O.uni     618  ----ACAC-- ---------- ----TGGG-C AATCTCTTGA AAATACTGAT
CS_O.uni    618  ----ACAC-- ---------- ----TGGG-C AATCTCTTGA AAATACTGAT
CA_O.uni    603  ----GGAC-- ---------- ----TGGG-C AATCTCTTCT AAATGCTGGT
M_O.uni(V)  603  ----GGAC-- ---------- ----TGGG-C AATCTCTTCT AAATGCCGGT
C_G.sac     651  ----AGAG-- ---------- ----AGGG-T AAGCCGTTCG AAATTCTGGT
CS_G.sac    651  ----AGAG-- ---------- ----AGGG-T AAGCCGTTCG AAATTCTGGT
C_G.rub,P   657  ----AGGG-- ---------- ----CGGG-T AACCTTTTCA AAATACCAGT
NA_G.rub,P  657  ----AGGG-- ---------- ----CGGG-T AACCTTTTCA AAATACCAGT
NA_G.rub,WI 646  ----AGGG-- ---------- ----CGGG-C AACCTTTTCA AAATACTGGT
CS_G.rub,WI 646  ----AGGG-- ---------- ----CGGG-C AACCTTTTCA AAATACTGGT
C_G.rub(P)  652  ----AGGG-- ---------- ----CGGG-C AACCTTTTCA AAATACCGGT
NA_G.rub,WII 645 ----AGGG-- ---------- ----CGGG-T AACCTCTTCG AAGTGCTGGC
C_G.rub,WII 645  ----AGGG-- ---------- ----CGGG-T AACCTCTTCG AAGTGCTGGC
C_G.con(P)  661  ----AGGG-- ---------- ----TGGG-T AACCTCTTCG AAGTGCTGGT
CS_G.con    662  ----AGGG-- ---------- ----TGGG-T AACCTCTTCG AAGTGCTGGT
NA_N.pac(D) 721  ----GGTG-- ---------- ACGTCGGG-T AAACATTCTC TAATGCTGAC
AA_N.pac(D) 721  ----GGTG-- ---------- ACGTCGGG-T AAACATTCTC TAATGCTGAC
C_N.dut     649  ----GTAA-- ---------- ---GTGGG-T AATCCATTGG AAGTAATGAT
CS_G.bul,Ia 618  ----ACTG-- ---------- ---GTCGT-T AGACTCGTGC AAGCAATTCA
CA_G.bul,IId 630 ----ACTG-- ---------- ---GTCGT-T AGACTCGTGC AATCCAATAG
M_G.bul(V)  625  ----ACTG-- ---------- ---GTCGT-T AGACTTTTAG TGCAATCCAT
Na_G.bul,Ib 627  ----ACTG-- ---------- ---GTCGT-T AGACTTTTAG TGCAATCCAT
NA_G.falc   595  ----ACTG-- ---------- ---GTCGT-T AGACTCGTGC GATCCCTCGG
CS_G.falc   596  ----ACTG-- ---------- ---GTCGT-T AGACTCGTGC GATCCCTCGG
Na_G.bul,IIb 627 ----ACTG-- ---------- ---GTCGT-T AGACTCGTGC AATCCAATAG
AA_G.bul,IIb 627 ----ACTG-- ---------- ---GTCGT-T AGACTCGTGC AATCCAATAG
Na-G.bul,IIa 625 ----ACTG-- ---------- ---GTCGT-T AGACTCGTGC AATCCAATAG
AA_G.bul,IIa 625 ----ACTG-- ---------- ---GTCGT-T AGACTCGTGC AATCCAATAG
AA_G.bul,IIc 628 ----ACTG-- ---------- ---GTCGT-T AGACTCGTGC AATCCAATAG
NA-T.qui,IIa 835 ----AAGG-- ---------- ---TTCCG-A CAACAGTCAA TCTAGTTATT
AA_T.qui,IIa 835 ----AAGG-- ---------- ---TTCCG-A CAACAGTCAA TCTAGTTATT
NA_T.qui,IIb 845 ----AAGG-- ---------- ---TTCCG-A CAACAGTCAA TCTAGTTATT
AA_T.qui,IIb 842 ----AAGG-- ---------- ---TTCCG-A CAACAGTCAA TCTAGTTATT
CS_T.qui,I  736  ----AGAG-- ---------- ---TTCCG-A CAACAGTCAA TCTAGTTATT
CS_G.glu    602  ----GCAA-- ---------- ---GTGGG-T AATCAATTAG AAGTAACGAT
NA_G.uvu    632  ----GCAA-- ---------- ---GTGGG-T AATCAATTAG AAGTAACGAT
A.becca_P   623  ----GTAA-- -AATTTCTCT ---GTGGG-T AATCCATTAG AAGTAATGAC
Troch_P     647  ----GTAA-- ---------- ---GTGGG-T AATCAATTAG AAGTAATGAT
H.germ_P    573  ----AGTT-- ---------- ---GTGGG-T AATCAATTTG AAGTAATGAT
Textu_P     635  ----GTAA-- ---------- ---GTGGG-T AATCAATTAG AAGTAATGAT
Biger_P     645  ----GTAA-- ---------- ---GTGGG-T AATCAATTAG AAGTAATGAT
Boliv_P     634  ----GTAA-- ---------- ---GCGGG-T AATCAATTAG AAGTAATGAT
G.oper_P    614  ----GTAA-- ---------- ---GTGGG-T AATCAATTAG AAGTAATGAT
A.trian_P   931  ----GTCA-- ---------- ---GTAGG-T AATCAATTAG AAGTAATGAT
As.rara_P   925  ----GTCA-- ---------- ---GTAGG-T AATCAATTAG AAGTAATGAT
A.angu_P    524  ----GTAA-- ---------- ---ATGGG-C AATCATTTAA AAATCGTGAT
E.acul_P    436  ----ATA--- ---------- ---GTTGG-T AATCA----G AAGTAATGAT
Quin_P      582  ----GTGA-- ---------- ---ATGGG-T AATCATTTAA AAATCGTGAT
M.secan_P   566  ----GTGA-- ---------- ---ATGGG-T AATCATTTGA AAATCGTGAT
P.pert_P    591  ----GTAA-- ---------- ---ATGGG-T AATCATTTAA AAATCGTGAT
Allogrom_P  589  ----GTAA-- ---------- ---GTTGG-T AATCAATTCG AAGTAATGAT
Align       318  ---------- ---------- ---------- ---------- ----------

C_G.sI      706  ----TCC--T CCT---CCCG TT---G--AG CA--TTTTA- --ATAATGG-
NA_G.sI     706  ----TCC--T CCT---CCCG TT---G--AG CA--TTTTA- --ATAATGG-
C_G.sI(V)   706  ----TCC--T CCT---CCCG TT---G--AG CA--TTTTA- --ATAATGG-
C_G.sIIa    689  ----TCC--T CCT---CCCG TT---G--AG CA--TTTTA- --ATAATGG-
```

```
   NA_G.sIIa   691  ----TCC--T CCT---CCCG TT---G--AG CA--TTTTA- --ATAATGG-
   CS_G.sIIa   678  ----TCC--T CCT---CCCG TT---G--AG CA--TTTTA- --ATAATGG-
   CA_G.sIIa   690  ----TCC--T CCT---CCCG TT---G--AG CA--TTTTA- --ATAATGG-
   CA_G.sIIb   683  ----TCC--T ACT---CCCG TT---G--AG CA--TTTTA- --ATAATGG-
   NA_G.sIIb   683  ----TCC--T ACT---CCCG TT---G--AG CA--TTTTA- --ATAATGG-
    NA_G.cal   703  ----TCC--T CTT---CCCG TT---G--AG CA--TTTTA- --ATAATGG-
      C_O.uni  647  ----AGA--A T--ACCCA-A CTCG---AGC AAATCTACTA CACGCTGGAG
     CS_O.uni  647  ----AGA--A T--ACCCA-A CTCG---AGC AAATCTACTA CACGCTGGAG
     CA_O.uni  632  ----AAG--G T--ACCCACA CTCG---AGC TATTTAATCA CACGCGGGAG
   M_O.uni(V)  632  ----AAG--G T--ACCCACA CTCG---AGC TATTTAATCA CACGCGGGAG
      C_G.sac  680  ----AAC-GA TT--CCCCGT AGTTA--AGC AAACTTAAAC CATAGTGGTG
     CS_G.sac  680  ----AAC-GA TT--CCCCGT AGTTA--AGC AAACTTAAAC CATAGTGGTG
   C_G.rub,P   686  ----TTG--A TTTT-A---- -A--AGC--- TTTGGGTCA- ATTAGGGTG-
  NA_G.rub,P   686  ----TTG--A TTTT-A---- -A--AGC--- TTTGGGTCA- ATTAGGGTG-
 NA_G.rub,WI   675  ----TTG--A TGTT-A---- -A--AGT--- ATTGAGTCG- TTTTTTTTGG
 CS_G.rub,WI   675  ----TTG--A TGTT-A---- -A--AGT--- ATTGAGTCG- TTTTTTTTGG
   C_G.rub(P)  681  ----TTG--A TATT-A---- -A--AGT--- ATAGAGTCG- TTTTTTTTGG
NA_G.rub,WII   674  ----ATG--A TTTCCCTAGT TTTGGCATTT GAATCCTTGT CGCAGCAGCC
 C_G.rub,WII   674  ----ATG--A TTTCCCTAGT TTTGGCATTT GAATCCTTGT CGCAGCAGCC
   C_G.con(P)  690  ----TTG--A TTTCCCCTAG TGTTTGTCAT AATAATCCTC GTCGCACCTA
     CS_G.con  691  ----TTG--A TTTCCC-TAG TGTTTGTCAT AATAATCCTC GTCGCACCTA
 NA_N.pac(D)   754  ----TTATTT TTCTGAAATA TGCACAACTC TAATATGACA TTCATTCCCG
 AA_N.pac(D)   754  ----TTATTT TTCTGAAATA TGCACAACTC TAATATGACA TTCATTCCCG
     C_N.dut   679  ----TTCTCT T--T-TTATA --------GC ACACCTAT-- -ATACGGCAT
  CS_G.bul,Ia  648  ----GAGCAA CGAATTG-CA AACACTCTTT GGGCACTGT- ---TTACTCC
CA_G.bul,IId   660  ----AAGAGG TGAATAAGTT TCCCAAAGTA CCGTGTTTAG CTCCCGTGTT
   M_G.bul(V)  655  ----TTGCAG CATCGAATCG CGAAAATGAG CATTGTTTAC ACCCGTTCTA
  Na_G.bul,Ib  657  ----TTGCAG CATCGAATCG CGAAAATGAG CATTGTTTAC ACCCGTTCTA
    NA_G.falc  625  ----AAAACG TGATTTCTGT TCCCAAACAA AGCACTTTCG TACCCTCGCT
    CS_G.falc  626  ----AAAACG TGATTTCTGT TCCCAAACAA AGCACTTTCG TACCCTCGCT
 Na_G.bul,IIb  657  ----AAGAGG TGAATAAGTT TCCCAAAGTA CCGTGTTTAG CTCCCGTGTT
  AA_G.bul,IIb 657  ----AAGAGG TGAATAAGTT TCCCAAAGTA CCGTGTTTAG CTCCCGTGTT
 Na-G.bul,IIa  655  ----AAGAGG TGACTAAGTT TCCCA--GTA CCGTGTTCGC CTCCCGTGTC
  AA_G.bul,IIa 655  ----AAGAGG TGACTAAGTT TCCCA--GTA CCGTGTTCGC CTCCCGTGTC
  AA_G.bul,IIc 658  ----AAGAGG TGACTCAGTA TCCCA--GTA CCGTGTTAGT CTCCCGTGTT
 NA-T.qui,IIa  865  ----GCTTGT AGTCGATTGC ATAATTGTAC AGTTCACGCA ACGCTCGGTT
  AA_T.qui,IIa 865  ----GCTTGT AGTCGATTGC ATAATTGTAC AGTTCACGCA ACGCTCGGTT
 NA_T.qui,IIb  875  ----GCTTGT AGTTGTTAAT TGTATAATTG TCTAGTTCAC GCAACGTTCG
  AA_T.qui,IIb 872  ----GCTTGT AGTTGTTAAT TGTATAATTG TCTAGTTCAC GCAACGTTCG
  CS_T.qui,I   766  ----GCTTGT AGTCGATCTC ATGATTGTTT AGTTTTATGC AACTTCGGGT
     CS_G.glu  632  ----TTCCCA AATTAGCACA CTTATATGCG GCGTTTGCGC CCGGGTACCA
     NA_G.uvu  662  ----TTCTTC GCATTAGCAC ACTTAATATA AGGCGTTGAC GCTCATGATT
    A.becca_P  662  ----TCGCAT AGACCATGGC ACACCAAATG TACGCGCAGG TCTACCCGGC
      Troch_P  677  ----TTCCTT TATTAATTTA TTAGT---GC ACAATTAT-- -ATACGGCAT
     H.germ_P  603  ----TTCTCC AAATATATAT ACTGCACACT CATATACAGT GTCTTATGTC
      Textu_P  665  ----TTCCTT TATATAGCAC ACATATATGC GGCATCTTTA CCCGGGTTAA
      Biger_P  675  ----TTCCTT TATATAGCAC ACATATATGC GGCATCTTTA CCCGGGTTAA
      Boliv_P  664  ----TTCCTA TTTTTCTCTG CACACATATA TGCGGCGTAT TTACCCGGCA
     G.oper_P  644  ----TTCCTT TATATGCACA TTTATGTTTG GCACTGATCC CCAGCCTAAC
    A.trian_P  961  ----TTCCCC TTTTTTATTT AAATGTATTT CAAA-TTTAT TTTTTTAATA
    As.rara_P  955  ----TTCCCC TTTTTTATTT AAATATAATA TAAAATTTAT TTTTTATTTG
     A.angu_P  554  ----TATTA- ---------- ---------- ---------- ----------
     E.acul_P  461  ----CTCTCT CGTATACGCA TACATGTATC GTATATATTA ATATATATTA
       Quin_P  612  ----TATTTC ATTTATACTA CATTTTATAG TTATAAATTT ATTAATATAC
    M.secan_P  596  ----TATTTC AATATACTAC ATTTTTAAGT TATAAATTTT TAATTATAGT
     P.pert_P  621  ----TATTTC ATACATATGG TATATTAATA TTGTATAATA CTATACATTT
   Allogrom_P  619  ----TTCCTT TT-------- --------GC ACAATAAT-- -AATATTTTA
        Align  322  ---------- ---------- ---------- ---------- ----------

      C_G.sI   736  -TC--TCTCT ACATCCCTAG CACATGATG- --TC--TAGT GCGATTG---
     NA_G.sI   736  -TC--TCTCT ACATCCCTAG CACATGATG- --TC--TAGT GCGATTG---
   C_G.sI(V)   736  -TC--TCTCT ACATCCCTAG CACATGATG- --TC--TAGT GCGATTG---
    C_G.sIIa   719  -CC--TCTCT ACATCCCTAG CA-AT-ATG- --CC--TAGT GCGATTG---
   NA_G.sIIa   721  -CC--TCTCT ACATCCCTAG CA-AT-ATG- --CC--TAGT GCGATTG---
   CS_G.sIIa   708  -CC--TCTCT ACATCCCTAG TA-AT-ATG- --AC--TAGT GCGATTG---
   CA_G.sIIa   720  -CC--TCTCT ACATCCCTAG CA-AT-ATG- --CC--TAGT GCGATTG---
   CA_G.sIIb   713  -CC--TCTCT ACATCCCTAG TA-AC-CCTA GTGCGATTGT AGTTGAGCCT
   NA_G.sIIb   713  -CC--TCTCT ACATCCCTAG TA-AC-CCTA GTGCGATTGT AGTTGAGCCT
    NA_G.cal   733  -TC--TCTCT ACGTCCCTAG TA-TATTCTA GTGCGATTGT AGTTGAG-TC
      C_O.uni  685  GCACAA-CGG -------TCC AAGGGCTTA- -TCCCTTGCG A------C--
     CS_O.uni  685  GCACAA-CGG -------TCC AAGGGCTTA- -TCCCTTGCG A------C--
     CA_O.uni  671  GCACACACGG -------TTC AAATCTCATT TGCGACTGTT GCCGCCATGA
   M_O.uni(V)  671  GCACACACGG -------TTC AAATCTCATT TGCGACTGTT GCCGCCATGA
      C_G.sac  721  TCA-AA-CGG GTCCGCTTCC ATCGGAAAAG ATTCCTCCGG AAAAAGGC--
     CS_G.sac  721  TCA-AA-CGG GTCCGCTTCC ATCGGAAAAG ATTCCTCCGG AAAAAGGC--
   C_G.rub,P   717  AT----GGTT CGATAG-AAC CAATATCA-C CCAGCTGCTA ------CTCG
  NA_G.rub,P   717  AT----GGTT CGATAG-AAC CAATATCA-C CCAGCTGCTA ------CTCG
```

```
NA_G.rub,WI    707  AT----GGTG GTTTGTTAAC CAATACCATC CGGCAACTGC -------TCG
CS_G.rub,WI    707  AT----GGTG GTTTGTTAAC CAATACCATC CGGCAACTGC -------TCG
 C_G.rub(P)    713  AT----GGTG GTTTACGAAC CAATACCATC CG-CAACTGC -------TCG
NA_G.rub,WII   718  CTGAGACGTA GTGAACGCTG CAAAGTGATT CCAGTCTTAG GTTTGC----
C_G.rub,WII    718  CTGAGACGTA GTGAACGCTG CAAAGTGATT CCAGTCTTAG GTTTGC----
 C_G.con(P)    734  CTCTAGGACA TAGTGAACAT TGTAGACTTT GTCATAATGA TTCCCGTCTT
   CS_G.con    734  CTCTAGGACA TAGTGAACAT TGTAGACTTT GTCATAATGA TTCCCGTCTT
NA_N.pac(D)    800  GGAGGACTAG TTGCCTCTTT GTGTGAGTGT AATGCAAACA TGAATAGCGA
AA_N.pac(D)    800  GGAGGACTAG TTGCCTCTTT GTGTGAGTGT AATGCAAACA TGAATAGCGA
   C_N.dut     711  TCATTCCCGG --GACGACTA GTTTCGTCTT --TTTTGTGC GAAT--GTAA
CS_G.bul,Ia    689  CGTTT--AC- AG--AGGAGA GATAGACTAA CCACCTATTC TCGACTCGC-
CA_G.bul,IId   706  CGTAGAGA-C CCCTCGCATT GCCAGTCGAT GTGGGTCCT- ----------
 M_G.bul(V)    701  GAGCAGTGGT TATTCTAACC ATATTTCCAC TGCTC----- ----------
Na_G.bul,Ib    703  GAGCAGTGGT TATTCTAACC ATATTTCCAC TGCTC----- ----------
  NA_G.falc    671  CAAACAAACG TGATTCCAAA TCCCTTACGA TTGCTACATC CTTTCTCGTA
  CS_G.falc    672  CAAACAAACT TTGAAACTCC TGGTAACTTT TTGGATTCTT TTCTTGTCAT
Na_G.bul,IIb   703  CGTAGCGACC CCTCGCAAAT GCCAGTCGAT GTGGGTCCT- ----------
 AA_G.bul,IIb  703  CGTAGCGACC CCTCGCAAAT GCCAGTCGAT GTGGGTCCT- ----------
Na-G.bul,IIa   699  AGTAGGG-CC CCTCGCATAT ACCAGTCGAA GTGGGCCCT- ----------
 AA_G.bul,IIa  699  AGTAGGG-CC CCTCGCATAT ACCAGTCGAA GTGGGCCCT- ----------
 AA_G.bul,IIc  702  CGTAGTGGCC CCGCGCATTG CCAGTCGTAG TGGGTCCT-- ----------
NA-T.qui,IIa   911  ATCATT---- ---------- ---------- ---------- ----------
 AA_T.qui,IIa  911  ATCATT---- ---------- ---------- ---------- ----------
NA_T.qui,IIb   921  TTTATCATT- ---------- ---------- ---------- ----------
 AA_T.qui,IIb  918  TTTATCATT- ---------- ---------- ---------- ----------
 CS_T.qui,I    812  CATGATTCA- ---------- ---------- ---------- ----------
   CS_G.glu    678  CTTGTTGGTA CTTCTGTGCG TGCAGATGTT TTTT-CCGCA TGT-------
   NA_G.uvu    708  CGCTCTCACG AGTGTTTCAT TGCACGTGTT GATGTGCGGG TTCGATTCCT
  A.becca_P    708  TCGCCTTTGT GTGAGTGCAG TGCGTAGCTT GTTGTTTCGT ACGTG-----
    Troch_P    717  CTTTACCCGG --CTTAAGCT TGCCTTAAG- --TTTTGTGC GTAT------
   H.germ_P    649  CATGAAAATT TTATTATTTT TGTGTGTGCA TTCGATGCTC G---------
    Textu_P    711  GCTGTCTTAA CTTTTGTGTG TATCGA---- ---------- ----------
    Biger_P    721  GCTTATCTTA ACTTCTGTGT GTATCGA--- ---------- ----------
    Boliv_P    710  TGCCTTGTTG CATGTTCTTT GTGCGTAGAT ATGTCCTCCG ----------
   G.oper_P    690  TCTTGTTAGT GCTTGTGTGC GTTCAGTGGA CCTTTTACGG TCTCTAA---
  A.trian_P   1006  TGTTTATTTA AAATTGATGC ACACTTTTAT GTCTATGTTT CTATT--AAC
  As.rara_P   1001  TGTTTATTTA AAATTGATGC ACACTTTTAT GTCTATGTTT CTATTTTTAC
   A.angu_P    559  ---------- ---------- ---------- ---------- ----------
   E.acul_P    507  ATTTATACGT A--------- ---------- ---------- ----------
     Quin_P    658  TTTATTGTAT TATTAATTAT TTATAA---- ---------- ----------
  M.secan_P    642  TTACTATTTT TATTATTTAT AA--------- ---------- ----------
   P.pert_P    667  TATAGTACTA TTA------- ---------- ---------- ----------
 Allogrom_P    646  TTCGATTAAT CTTAGTCTTT TTAGATTTTG TATTAAAGTT AAAA------
      Align    322  ---------- ---------- ---------- ---------- ----------

    C_G.sI     775  TAGTTGAGTC TTG--CCATT ---------- ---------- ----------
   NA_G.sI     775  TAGTTGAGTC TTG--CCATT ---------- ---------- ----------
  C_G.sI(V)    775  TAGTTGAGTC TTG--CCATT ---------- ---------- ----------
   C_G.sIIa    756  TAGTTGAGCC TTG--CCATT ---------- ---------- ----------
  NA_G.sIIa    758  TAGTTGAGCC TTG--CCATT ---------- ---------- ----------
  CS_G.sIIa    745  TAGTTGAGCC TTG--CCATT ---------- ---------- ----------
  CA_G.sIIa    757  TAGTTGAGCC TTG--CCATT ---------- ---------- ----------
  CA_G.sIIb    758  TGCCATT--- ---------- ---------- ---------- ----------
  NA_G.sIIb    758  TGCCATT--- ---------- ---------- ---------- ----------
   NA_G.cal    778  TGCCATT--- ---------- ---------- ---------- ----------
    C_O.uni    717  TGGTGCAGCC AATG---ACG ---------- ---------- ----------
   CS_O.uni    717  TGGTGCAGCC AATG---ACG ---------- ---------- ----------
   CA_O.uni    714  ---------- ---------- ---------- ---------- ----------
  M_O.uni(V)   714  ---------- ---------- ---------- ---------- ----------
    C_G.sac    767  TTATGCAGGC ATT-TC-ACG ---------- ---------- ----------
   CS_G.sac    767  TTATGCAGGC ATT-TC-ACG ---------- ---------- ----------
  C_G.rub,P    755  TGCGCGCGTG -T--------- ---------- ---------- ----------
 NA_G.rub,P    755  TGCGCGCGTG -T--------- ---------- ---------- ----------
NA_G.rub,WI    746  CGAGTTTGTG GTT------- ---------- ---------- ----------
CS_G.rub,WI    746  CGAGTTTGTG GTT------- ---------- ---------- ----------
 C_G.rub(P)    751  CGAGTTTGTG GTT------- ---------- ---------- ----------
NA_G.rub,WII   764  ---------- ---------- ---------- ---------- ----------
C_G.rub,WII    764  ---------- ---------- ---------- ---------- ----------
 C_G.con(P)    784  AGGTTTGCGA CGCATGCTTT ---------- ---------- ----------
   CS_G.con    784  AGGTTTGCGA CGCATGCTTT ---------- ---------- ----------
NA_N.pac(D)    850  CTGTCGCTGT TTACTCA--- ---------- ---------- ----------
AA_N.pac(D)    850  CTGTCGCTGT TTACTCA--- ---------- ---------- ----------
   C_N.dut     755  TGTATTCTTT ATCCG----- ---------- ---------- ----------
CS_G.bul,Ia    733  ---------- ---------- ---------- ---------- ----------
CA_G.bul,IId   744  ---------- ---------- ---------- ---------- ----------
 M_G.bul(V)    736  ---------- ---------- ---------- ---------- ----------
Na_G.bul,Ib    738  ---------- ---------- ---------- ---------- ----------
```

```
   NA_G.falc   721  ACGTCTCGCC TAGCGGAGCT TTCAACTTT- ---------- ----------
   CS_G.falc   722  CTTTTCGTGT CATCTTT--- ---------- ---------- ----------
Na_G.bul,IIb   742  ---------- ---------- ---------- ---------- ----------
  AA_G.bul,IIb 742  ---------- ---------- ---------- ---------- ----------
Na-G.bul,IIa   737  ---------- ---------- ---------- ---------- ----------
  AA_G.bul,IIa 737  ---------- ---------- ---------- ---------- ----------
  AA_G.bul,IIc 740  ---------- ---------- ---------- ---------- ----------
NA-T.qui,IIa   917  ---------- ---------- ---------- ---------- ----------
  AA_T.qui,IIa 917  ---------- ---------- ---------- ---------- ----------
NA_T.qui,IIb   930  ---------- ---------- ---------- ---------- ----------
  AA_T.qui,IIb 927  ---------- ---------- ---------- ---------- ----------
  CS_T.qui,I   821  ---------- ---------- ---------- ---------- ----------
    CS_G.glu   720  ---------- ---------- ---------- ---------- ----------
    NA_G.uvu   758  GCTTTACCTT ATGT------ ---------- ---------- ----------
   A.becca_P   753  ---------- ---------- ---------- ---------- ----------
     Troch_P   756  -CGATGTTTT TTCCG----- ---------- ---------- ----------
     H.germ_P  690  ---------- ---------- ---------- ---------- ----------
     Textu_P   737  ----TGTTTT TTCCG----- ---------- ---------- ----------
     Biger_P   748  ----TGTTTT TTCCG----- ---------- ---------- ----------
     Boliv_P   750  ---------- ---------- ---------- ---------- ----------
     G.oper_P  737  ---------- ---------- ---------- ---------- ----------
   A.trian_P  1054  ATATTAGGAT ATTAATACTA TTTATTAATA GTTTAATTTC TAATTTTGTT
   As.rara_P  1051  ATGTTAGAAT ATTAATACTA TTTATTAATA GTTTAATTTT TAATTTTGTG
    A.angu_P  559   ---------- ---------- ---------- ---------- ----------
    E.acul_P  518   ---------- ---------- ---------- ---------- ----------
      Quin_P  684   ---------- ---------- ---------- ---------- ----------
   M.secan_P  664   ---------- ---------- ---------- ---------- ----------
    P.pert_P  680   ---------- ---------- ---------- ---------- ----------
  Allogrom_P  690   ---------- ---------- ---------- ---------- ----------
       Align  322   ---------- ---------- ---------- ---------- ----------

      C_G.sI  793   ---------- ---------- ---------- ---------- ----------
     NA_G.sI  793   ---------- ---------- ---------- ---------- ----------
    C_G.sI(V) 793   ---------- ---------- ---------- ---------- ----------
    C_G.sIIa  774   ---------- ---------- ---------- ---------- ----------
    NA_G.sIIa 776   ---------- ---------- ---------- ---------- ----------
    CS_G.sIIa 763   ---------- ---------- ---------- ---------- ----------
    CA_G.sIIa 775   ---------- ---------- ---------- ---------- ----------
    CA_G.sIIb 765   ---------- ---------- ---------- ---------- ----------
    NA_G.sIIb 765   ---------- ---------- ---------- ---------- ----------
     NA_G.cal 785   ---------- ---------- ---------- ---------- ----------
      C_O.uni 734   ---------- ---------- ---------- ---------- ----------
     CS_O.uni 734   ---------- ---------- ---------- ---------- ----------
     CA_O.uni 714   ---------- ---------- ---------- ---------- ----------
   M_O.uni(V) 714   ---------- ---------- ---------- ---------- ----------
      C_G.sac 785   ---------- ---------- ---------- ---------- ----------
     CS_G.sac 785   ---------- ---------- ---------- ---------- ----------
    C_G.rub,P 766   ---------- ---------- ---------- ---------- ----------
   NA_G.rub,P 766   ---------- ---------- ---------- ---------- ----------
  NA_G.rub,WI 759   ---------- ---------- ---------- ---------- ----------
  CS_G.rub,WI 759   ---------- ---------- ---------- ---------- ----------
   C_G.rub(P) 764   ---------- ---------- ---------- ---------- ----------
 NA_G.rub,WII 764   ---------- ---------- ---------- ---------- --------GA
  C_G.rub,WII 764   ---------- ---------- ---------- ---------- --------GA
   C_G.con(P) 804   ---------- ---------- ---------- ---------- ----------
     CS_G.con 804   ---------- ---------- ---------- ---------- ----------
  NA_N.pac(D) 867   ---------- ---------- ---------- ---------- ----------
  AA_N.pac(D) 867   ---------- ---------- ---------- ---------- ----------
      C_N.dut 770   ---------- ---------- ---------- ---------- ----------
  CS_G.bul,Ia 733   ---------- ---------- ---------- ---------- ----------
 CA_G.bul,IId 744   ---------- ---------- ---------- ---------- ----------
   M_G.bul(V) 736   ---------- ---------- ---------- ---------- ----------
  Na_G.bul,Ib 738   ---------- ---------- ---------- ---------- ----------
    NA_G.falc 750   ---------- ---------- ---------- ---------- ----------
    CS_G.falc 739   ---------- ---------- ---------- ---------- ----------
Na_G.bul,IIb  742   ---------- ---------- ---------- ---------- ----------
  AA_G.bul,IIb 742  ---------- ---------- ---------- ---------- ----------
Na-G.bul,IIa  737   ---------- ---------- ---------- ---------- ----------
  AA_G.bul,IIa 737  ---------- ---------- ---------- ---------- ----------
  AA_G.bul,IIc 740  ---------- ---------- ---------- ---------- ----------
NA-T.qui,IIa  917   ---------- ---------- ---------- ---------- ----------
  AA_T.qui,IIa 917  ---------- ---------- ---------- ---------- ----------
NA_T.qui,IIb  930   ---------- ---------- ---------- ---------- ----------
  AA_T.qui,IIb 927  ---------- ---------- ---------- ---------- ----------
  CS_T.qui,I  821   ---------- ---------- ---------- ---------- ----------
    CS_G.glu  720   ---------- ---------- ---------- ---------- ----------
    NA_G.uvu  772   ---------- ---------- ---------- ---------- ----------
```

```
A.becca_P   753  ---------- ---------- ---------- ---------- ----------
  Troch_P   770  ---------- ---------- ---------- ---------- ----------
 H.germ_P   690  ---------- ---------- ---------- ---------- ----------
  Textu_P   748  ---------- ---------- ---------- ---------- ----------
  Biger_P   759  ---------- ---------- ---------- ---------- ----------
  Boliv_P   750  ---------- ---------- ---------- ---------- ----------
 G.oper_P   737  ---------- ---------- ---------- ---------- ----------
A.trian_P  1104  ATAGTTACTG ACATGTGCTC TCATGTTTTA T---------- ----------
As.rara_P  1101  ATAGTTACTG ACATGTGCTC TCATATTT-- ---------- ----------
  A.angu_P   559  ---------- ---------- ---------- ---------- ----------
  E.acul_P   518  ---------- ---------- ---------- ---------- ----------
    Quin_P   684  ---------- ---------- ---------- ---------- ----------
M.secan_P   664  ---------- ---------- ---------- ---------- ----------
 P.pert_P   680  ---------- ---------- ---------- ---------- ----------
Allogrom_P  690  ---------- ---------- ---------- ---------- ----------
     Align  322  ---------- ---------- ---------- ---------- ----------

   C_G.sI   793  ---TATGC-- AA---GTGTC AATT-----C TCA--GTG-G GGACAGCCAT
  NA_G.sI   793  ---TATGC-- AA---GTGTC AATT-----C TCA--GTG-G GGACAGCCAT
C_G.sI(V)   793  ---TATGC-- AA---GTGTC AATT-----C TCA--GTG-G GGACAGCCAT
  C_G.sIIa  774  ---TATGC-- AA---GTGTC AATT-----C TCA--GTG-G GGACAGACAT
 NA_G.sIIa  776  ---TATGC-- AA---GTGTC AATT-----C TCA--GTG-G GGACAGACAT
 CS_G.sIIa  763  ---TATGC-- AA---GTGTC AATT-----C TCA--GTG-G GGACAGACAT
 CA_G.sIIa  775  ---TATGC-- AA---GTGTC AATT-----C TCA--GTG-G GGACAGACAT
 CA_G.sIIb  765  ---TATGC-- AA---GTGTC AATT-----C TCA--GTG-G GGACAGACAT
 NA_G.sIIb  765  ---TATGC-- AA---GTGTC AATT-----C TCA--GTG-G GGACAGACAT
  NA_G.cal  785  ---TATGC-- AA---GTGTC AATT-----C TCA--GTG-G GGACAGACAT
   C_O.uni  734  ---TATGC-- AAC--TTGTC AACT-----C TAC--GTG-G GGATAGTCGC
  CS_O.uni  734  ---TATGC-- AAC--TTGTC AACT-----C TAC--GTG-G GGATAGTCGC
  CA_O.uni  714  -CGTATGC-- AAC--TTGTC AATT-----C TTC--GTG-G GGATCGTCGA
M_O.uni(V)  714  -CGTATGC-- GAC--TTGTC AATT-----C TTC--GTG-G GGATCGTCGA
   C_G.sac  785  ---TATGC-- -TC--CTATA AATT-----C CTG--GTA-G GGATAGTCTA
  CS_G.sac  785  ---TATGC-- -TC--CTATA AATT-----C CTG--GTA-G GGATAGTCTA
 C_G.rub,P  766  ---TATTC-- -----CTGGT GACT-----C ATG--GTG-G GGACCGACAT
NA_G.rub,P  766  ---TATTC-- -----CTGGT GACT-----C ATG--GTG-G GGACCGACAT
NA_G.rub,WI 759  ---------- -----CTGGC GACT-----C ATG--GTG-G GGACCGACGT
CS_G.rub,WI 759  ---------- -----CTGGC GACT-----C ATG--GTG-G GGACCGACGT
 C_G.rub(P)  764  ---------- -----CTGGT GACT-----C ATG--GTG-G GGACCGATGT
NA_G.rub,WII 766 CACATGCAAT TCC--TGGTT GACT-----C ATC--GTG-G GGACTGATTC
C_G.rub,WII 766  CACATGCAAT TCC--TGGTT GACT-----C ATC--GTG-G GGACTGATTC
 C_G.con(P)  804  ---------- TCC--TGGTT GACT-----C ATC--GTG-G GAACTGATTC
  CS_G.con   804  ---------- TCC--TGGTT GACT-----C ATC--GTG-G GAACTGATTC
NA_N.pac(D) 867  ---TATGT-G CCC--CTGTC AATT-----C GTG--GTG-G GGACAGACCA
AA_N.pac(D) 867  ---TATGT-G CCC--CTGTC AATT-----C GTG--GTG-G GGACAGACCA
   C_N.dut  770  ---TATGT-G CGA--TTGTC AATT-----C ATG--GTG-G GGACAGACCA
CS_G.bul,Ia 733  ---CCGCC-- CAT--CTTTC AATT-----C TTG--GTG-G GGACAGTAGG
CA_G.bul,IId 744  ---TTGTG-- CTT--TTTTC AAAT-----C TTA--GTG-G GGACAGACAT
 M_G.bul(V)  736  ---GCCTC-T CAA--CTGTC AATT-----C TTG--GTG-G GGACAGTCAG
Na_G.bul,Ib 738  ---GCCTC-T CAA--CTGTC AATT-----C TTG--GTG-G GGACAGTCAG
  NA_G.falc  750  ---TCGTC-G CAA--CTCTC AATT-----C ACA--GTG-G GGACAGTCGT
  CS_G.falc  739  ---TCGTC-G CAA--CTCTC AATT-----C ACA--GTG-G GGACAGTCGT
Na_G.bul,IIb 742  ---TTGTG-- CTT--TTTTC AATT-----C TTA--GTG-G GGACAGACAT
AA_G.bul,IIb 742  ---TTGTG-- CTT--TTTTC AATT-----C TTA--GTG-G GGACAGACAT
Na-G.bul,IIa 737  ---TTGTG-- CTT--TTTTC AAAT-----C TTA--GTG-G GGACAGACAT
AA_G.bul,IIa 737  ---TTGTG-- CTT--TTTTC AAAT-----C TTA--GTG-G GGACAGACAT
AA_G.bul,IIc 740  ---TTGTG-- CTT--TTTTC AAAT-----C TTA--GTG-G GGACAGACAT
NA-T.qui,IIa 917  ---CGCTA-- GTA--TTGTT AATT-----C ACA--GTG-G GGACAGTCGT
AA_T.qui,IIa 917  ---CGCTA-- GTA--TTGTT AATT-----C ACA--GTG-G GGACAGTCGT
NA_T.qui,IIb 930  ---CTCTA-- GTA--TTGTT AATT-----C ACA--GTG-G GGACAGTCGT
AA_T.qui,IIb 927  ---CTCTA-- GTA--TTGTT AATT-----C ACA--GTG-G GGACAGTCGT
 CS_T.qui,I  821  ---CTAGT-- ACG--CTTCT AATT-----C ACA--GTG-G GGACAGTCGT
   CS_G.glu  720  ---------G CAA--TTGTC AATT-----C ATG--GTG-G GGACAGACCA
   NA_G.uvu  772  ---------G CAA--TTGTC AATT-----C ATG--GTG-G GGACAGACCA
  A.becca_P  753  -----CCA-C TCC--GTATT AATT-----C GTAC-GTG-G GGATAGATCA
    Troch_P  770  ---TATGT-G CAA--TTGTC AATT-----C ATG--GTG-G GGACAGACCA
   H.germ_P  690  ---TATGT-G CAA--TTGTC AATT-----C ATG--GTG-G GGACAGACCA
    Textu_P  748  ---CATGT-G CAA--TTGTC AATT-----C ATG--GTG-G GGACAGACCA
    Biger_P  759  ---CATGT-G CAA--TTGTC AATT-----C ATG--GTG-G GGACAGACCA
    Boliv_P  750  ---CATGT-G CAA--TTGTC AATT-----C ATG--GTG-G GGACAGACCA
   G.oper_P  737  CCATTCGT-G CAA--TTGTC AATT-----C ATG--GTG-G GGACAGACCA
  A.trian_P 1135  ---------T AAA--TGTTC AATT-----C GTG--GTG-G GGACAGACCA
  As.rara_P 1129  ---------T AAA--TGTTC AATT-----C GTG--GTG-G GGACAGACCA
   A.angu_P  559  ---TAATA-C ACA--TATTT AATT-----T AAG--GTG-G GGATAGTGTA
   E.acul_P  518  ---TACGC-G TAA--ATATT AATT-----C ATG--GTG-G GGATAGTCCA
     Quin_P  684  ---CTGTA-G TAT--TAATT AATT-----T AAG--GTG-G GGATAGTGAA
  M.secan_P  664  ---CTGTA-G TAT--TAATT AATT-----T AAG--GTG-G GGATAGTGAA
   P.pert_P  680  --CAAATA-C CAT--TAATT AATT-----T AAG--GTG-G GGATAGTGTA
```

```
Allogrom_P   690  ---TATGT-G CT---CCTTT ATTT-----C ATG--GTG-G GGACTGACCA
     Align   322  ---------- --------mm mmmm-----m mmm--mmm-m mmmmmmmmmm

     C_G.sI  827  TTGA--TAAT T-CTTTGGCT CGGCCTCAAC TAGGAATGCC TTGTACGG-G
    NA_G.sI  827  TTGA--TAAT T-CTTTGGCT CGGCCTCAAC TAGGAATGCC TTGTACGG-G
  C_G.sI(V)  827  TTGA--TAAT T-CTTTGGCT CGGCCTCAAC TAGGAATGCC TTGTACGG-G
   C_G.sIIa  808  TTGA--TAAT T-CTTTGTCT CGTTCTTAAC TAGGAATGCC TTGTACGG-G
  NA_G.sIIa  810  TTGA--TAAT T-CTTTGTCT CGTTCTTAAC TAGGAATGCC TTGTACGG-G
  CS_G.sIIa  797  TTGA--TAAT T-CTTTGTCT CGTTCTTAAC TAGGAATGCC TTGTACGG-G
  CA_G.sIIa  809  TTGA--TAAT T-CTTTGTCT CGTTCTTAAC TAGGAATGCC TTGTACGG-G
  CA_G.sIIb  799  TTGA--TAAT T-CTTTGTCT CGTTCTTAAC TAGGAATGCC TTGTACGG-G
  NA_G.sIIb  799  TTGA--TAAT T-CTTTGTCT CGTTCTTAAC TAGGAATGCC TTGTACGG-G
   NA_G.cal  819  TTGA--TAAT T-CTTTGTCT CGTTCTTAAC TAGGAATGCC TTGTACGG-G
    C_O.uni  769  CTG---AAAT T-CTTCGACT CG-TTCTAAC TAGGAATGCC TTGTACGG-G
   CS_O.uni  769  CTG---AAAT T-CTTCGACT CG-TTCTAAC TAGGAATGCC TTGTACGG-G
   CA_O.uni  751  TTG---CAAC T-ATTCGACT CG-CCACAAC TAGGAATGCC TCGTACGG-G
 M_O.uni(V)  751  TTG---CAAC T-ATTCGACT CG-CCACAAC TAGGAATGCC TCGTACGG-G
    C_G.sac  819  TTG---TAAC T-GGTAGACT TG---TCAAC TAGGAATGCC TTGTACGG-G
   CS_G.sac  819  TTG---TAAC T-GGTAGACT TG---TCAAC TAGGAATGCC TTGTACGG-G
  C_G.rub,P  798  TTG---TAAT T-GTCTGTCG CG-TGTTAAC CGGGAATGCC TTGTACTG-T
 NA_G.rub,P  798  TTG---TAAT T-GTCTGTCG CG-TGTTAAC CGGGAATGCC TTGTACTG-T
NA_G.rub,WI  786  TTG---TAAT T-TTTTGTCG CG-TGTTAAC TAGGAATGCC TTGTACTG-T
CS_G.rub,WI  786  TTG---TAAT T-TTTTGTCG CG-TGTTAAC TAGGAATGCC TTGTACTG-T
  C_G.rub(P)  791  TTG---TAAT T-GTTTGTCG CG-TGTCAAC TAGGAATGCC TTGTACTG-T
NA_G.rub,WII  806  TTG---TAAT T-ATTTTTCA CGG-TTCAAC CAGGAATGCC TTGTACCG-G
 C_G.rub,WII  806  TTG---TAAT T-ATTTTTCA CGG-TTCAAC CAGGAATGCC TTGTACCG-G
  C_G.con(P)  834  TTG---TAAT T-ATTTGTCA CGG-TTCAAC CAGGAATGCC TTGTACTG-G
    CS_G.con  834  TTG---TAAT T-ATTTGTCA CGG-TTCAAC CAGGAATGCC TTGTACTG-G
 NA_N.pac(D)  903  TTGC--TAAT T-GTTGGTCT CGCTTATAAC TAGGAATGCC TTGTACGG-G
 AA_N.pac(D)  903  TTGC--TAAT T-GTTGGTCT CGCTTATAAC TAGGAATGCC TTGTACGG-G
    C_N.dut   806  TTGT--TAAT T-GTTGGTCT CGGTCTTAAC TAGGAATGCC TTGTACGG-G
 CS_G.bul,Ia  768  TTGT--TAAC T-TTCTTACT CGGTCCTAAC CAGGAATGCC TCGTACAG-G
CA_G.bul,IId  779  CTGT--TAAC T-TTTTGTCT CGGTCCCAAC CAGGAATGCC TCGTACAG-G
  M_G.bul(V)  772  TTGT--TAAC T-TTCTGACT CG-TCTGAAC CAGGAATGCC TCGTACAG-G
 Na_G.bul,Ib  774  TTGT--TAAC T-TTCTGACT CG-TCTGAAC CAGGAATGCC TCGTACAG-G
   NA_G.falc  786  TTGT--CAAC T-TTTCGTCT CGGTCCCAAC TAGGAATGCC TCGTACGA-G
   CS_G.falc  775  TTGT--CAAC T-TTTCGTCT CGGTCTCAAC TAGGAATGCC TCGTACGA-G
Na_G.bul,IIb  777  CTGT--TAAC T-TTTTGTCT CGGTCCCAAC CAGGAATGCC TCGTACAG-G
 AA_G.bul,IIb 777  CTGT--TAAC T-TTTTGTCT CGGTCCCAAC CAGGAATGCC TCGTACAG-G
Na-G.bul,IIa  772  CTGT--TAAC T-TTTTGTCT CGGTCCCAAC CAGGAATGCC TCGTACAG-G
 AA_G.bul,IIa 772  CTGT--TAAC T-TTTTGTCT CGGTCCCAAC CAGGAATGCC TCGTACAG-G
 AA_G.bul,IIc 775  CTGT--TAAC T-TTTTGTCT CGGTCCCAAC CAGGAATGCC TCGTACAG-G
NA-T.qui,IIa  952  TTG---TAAT T-CTGAGACT CGGT-TCAAC CAGGAATGCC TCGTATTT-G
 AA_T.qui,IIa 952  TTG---TAAT T-CTGAGACT CGGT-TCAAC CAGGAATGCC TCGTATTT-G
NA_T.qui,IIb  965  TTG---TAAT T-CTGAGACT CGGT-TCAAC CAGGAATGCC TCGTATTT-G
 AA_T.qui,IIb 962  TTG---TAAT T-CTGAGACT CGGT-TCAAC CAGGAATGCC TCGTATTT-G
 CS_T.qui,I   856  TTG---TAAT T-CTGAGACT CGGT-TCAAC TAGGAATGCC TAGTATTG-A
    CS_G.glu  751  TTGT--TAAT T-GTTGGTCT CGGTCTTAAC TAGGAATGCC TTGTACGG-G
    NA_G.uvu  803  TTGT--TAAT T-GTTGGTCT CGGTCTTAAC TAGGAATGCC TTGTATGA-G
   A.becca_P  788  TTGTT-TAAT T-GTTGGTCT CGGTCTTAAC TAGGAATGCC TTGTACGG-G
     Troch_P  806  TTGT--TAAG T-GTTGGTCT CGGTCTTAAC TAGGAATGCC TTGTACGG-G
    H.germ_P  726  TTGT--TAAT T-GTTGGTCT CGGTCTTAAC TAGGAATGCC TTGTACGG-G
     Textu_P  784  TTGT--TAAT T-GTTGGTCT CGGTCTTAAC TAGGAATGCC TTGTACGG-G
     Biger_P  795  TTGT--TAAT T-GTTGGTCT CGGTCTTAAC TAGGAATGCC TTGTACGG-G
     Boliv_P  786  TTGTTTCAAT T-GTTGGTCT CGGTCTTAAC TAGGAATGCC TTGTACGG-G
    G.oper_P  776  TTGT--TAAT T-GTTGGTCT CGGTCTTAAC TAGGAATGCC TTGTACGG-G
   A.trian_P 1166  TTGT--TAAT T-GTTGGTCT CGGTCTCAAC TAGGAATGCC TTGTACGG-G
   As.rara_P 1160  TTGT--TAAT T-GTTGGTCT CGGTCTCAAC TAGGAATGCC TTGTACGG-G
    A.angu_P  595  TTGT--TAAT T-ATTACACT TGGCCTTAAC TAGGAATGCC TTGTACTC-T
    E.acul_P  554  TTGT--TAAT T-GTTGGTCT CGGTCTAAAC TAGGAATGCC TTGTACGG-G
      Quin_P  720  TTGT--TAAT T-ATTTCACT TGGCCTTAAC TAGGAATGCC TTGTACTG-T
   M.secan_P  700  TTGT--TAAT T-ATTTCACT TGGCCTTAAC TAGGAATGCC TTGTACTG-T
    P.pert_P  717  TTGT--TAAT T-ATTACACT TGGCCTTAAC TAGGAATGCC TTGTACTC-T
  Allogrom_P  725  TTGT--TAAT T-GTTGGTCA CG-TCTCAAC TAGGAATGCC TTGTACTG-G
     Align    346  mmm---mmmm m-mmmmmmmmm mm---mmmmm mmmmmmmmmm mmmmmmmmm-m

     C_G.sI   873  TCTT-GGTTC AAC-AGACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
    NA_G.sI   873  TCTT-GGTTC AAC-AGACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
  C_G.sI(V)   873  TCTT-GGTTC AAC-AGACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
   C_G.sIIa   854  TCTT-GGTTC AAC-AGACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
  NA_G.sIIa   856  TCTT-GGTTC AAC-AGACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
  CS_G.sIIa   843  TCTT-GGTTC AAC-AGACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
  CA_G.sIIa   855  TCTT-GGTTC AAC-AGACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
  CA_G.sIIb   845  TCTT-GGTTC AAC-AGACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
  NA_G.sIIb   845  TCTT-GGTTC AAC-AGACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
   NA_G.cal   865  TCTT-GGTTC AAC-AGACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
    C_O.uni   813  CCTT-GGTTC ATT-ATGCCG CCCGGAATAC GTCCCTGCCC TTTGTACACA
```

```
   CS_O.uni   813  CCTT-GGTTC ATT-ATGCCG CCCGGAATAC GTCCCTGCCC TTTGTACACA
   CA_O.uni   795  TCTC-GGTTC ACC-ATACCG CCCGGAATAC GTCCCTGCCC TTTGTACACA
  M_O.uni(V)  795  TCTC-GGTTC ACC-ATACCG CCCGGAATAC GTCCCTGCCC TTTGTACACA
    C_G.sac   861  TG--GGGTAG CAC-ATTCCA CCCGGGATAC GTCCCTGCCT TTTGTACACA
   CS_G.sac   861  TG--GGGTAG CAC-ATTCCA CCCGGGATAC GTCCCTGCCT TTTGTACACA
  C_G.rub,P   842  CG----ATTC ACT-AAGTTA TGGGGAATAC GTCCCTGCCC TTTGTACACA
 NA_G.rub,P   842  CG----ATTC ACT-AAGTTA TGGGGAATAC GTCCCTGCCC TTTGTACACA
 NA_G.rub,WI  830  CG----ATTC ACT-AAGTTA CTGGGAATAC GTCCCTGCCC TTTGTACACA
 CS_G.rub,WI  830  CG----ATTC ACT-AAGTTA CTGGGAATAC GTCCCTGCCC TTTGTACACA
  C_G.rub(P)  835  CG----ATTC ACT-AAGTTA TTGGGAATAC GTCCCTGCCC TTTGTACACA
NA_G.rub,WII  850  CG----GCTC ATT-AAACCG CTGGGAATAC GTCCCTGCCC TTTGTACACA
 C_G.rub,WII  850  CG----GCTC ATT-AAACCG CTGGGAATAC GTCCCTGCCC TTTGTACACA
  C_G.con(P)  878  CG----GCTC ATT-AAACCG CTCGGTATAC GTCCCTGCCC TTTGTACACA
   CS_G.con   878  CG----GCTC ATT-AAACCG CTCGGTATAC GTCCCTGCCC TTTGTACACA
 NA_N.pac(D)  949  TCTTTGGTTC AAC-AAACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
 AA_N.pac(D)  949  TCTTTGGTTC AAC-AAACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
    C_N.dut   852  TCTTTGGTTC AAC-AAACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
 CS_G.bul,Ia  814  TT---GGTTC ACC-ATACCA CCTGGAATTA GTCCCTGCCC TTTGTACACA
CA_G.bul,IId  825  TT---GGTTC ACC-ATACCA CCTGGAATTA GTCCCTGCCC TTTGTACACA
  M_G.bul(V)  817  TT---GGTTC ACC-ATACCA CCTGGAATA- GTCCCTGCCC TTTGTACACA
 Na_G.bul,Ib  819  TT---GGTTC ACC-ATACCA CCTGGAATTA GTCCCTGCCC TTTGTACACA
   NA_G.falc  832  TT---GGTTC ATC-AGACCA CTCGGAATAC GTCCCTGCCC TTTGTACACA
   CS_G.falc  821  TT---GGTTC ATC-AGACCA CTCGGAATAC GTCCCTGCCC TTTGTACACA
Na_G.bul,IIb  823  TT---GGTTC ACC-ATACCA CCTGGAATTA GTCCCTGCCC TTTGTACACA
 AA_G.bul,IIb 823  TT---GGTTC ACC-ATACCA CCTGGAATTA GTCCCTGCCC TTTGTACACA
Na-G.bul,IIa  818  TT---GGTTC ACC-ATACCA CCTGGAATTA GTCCCTGCCC TTTGTACACA
 AA_G.bul,IIa 818  TT---GGTTC ACC-ATACCA CCTGGAATTA GTCCCTGCCC TTTGTACACA
 AA_G.bul,IIc 821  TT---GGTTC ACC-ATACCA CCTGGAATTA GTCCCTGCCC TTTGTACACA
NA-T.qui,IIa  996  TG----GTTC ATT-ATACTC CGTGGAATAA GTCCCTGTCC TTTGTACACA
 AA_T.qui,IIa 996  TG----GTTC ATT-ATACTC CGTGGAATAA GTCCCTGTCC TTTGTACACA
NA_T.qui,IIb 1009  TG----GTTC ATT-AAACTC CGTGGAATAA GTCCCTGTCC TTTGTACACA
 AA_T.qui,IIb 1006 TG----GTTC ATT-AAACTC CGTGGAATAA GTCCCTGTCC TTTGTACACA
  CS_T.qui,I  900  TG----GTTC ACT-AAACTT TCTGGAATAA GTCCCTGCCC TTTGTACACA
    CS_G.glu  797  TCTTTGGTTC AAC-AAACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
    NA_G.uvu  849  TCTTTGGTTC AAC-AAACCA CTCAGAATAT GTCCCTGCCC TTTGTACACA
   A.becca_P  835  TCTCTGGTTC AAC-ATACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
     Troch_P  852  TCTTTGGTTC AAC-AAACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
    H.germ_P  772  TCTTTGGTTC AAC-AAACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
     Textu_P  830  TCTTTGGTTC AAC-AAACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
     Biger_P  841  TCTTTGGTTC AAC-AAACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
     Boliv_P  834  TCTTTGGTTC AAC-AAACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
    G.oper_P  822  TCTTTGGTTC AAC--AACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
   A.trian_P 1212  TCTTTGGTTC AAC-AAACCA CCCGGAATAT GTCCCTGCCC TTTGTACACA
   As.rara_P 1206  TCTTTGGTTC AAC-AAACCA CCCGGAATAT GTCCCTGCCC TTTGTACACA
    A.angu_P  641  TCTTTGGTTT AAC-ATACCA AGAGGAATAC GTCCCTGCCC TTTGTACACA
    E.acul_P  600  TTATTGGTTC AAC-AAACCA CCCGGAATAC GTCCCTGCCC TTTGTACACA
      Quin_P  766  TCCTTGGTTC AAC-ATACCA ACAGGAATAC GTCCCTGCCC TTTGTACACA
   M.secan_P  746  TCCTTGGTTC AAC-ATACCA ACAGGAATAC GTCCCTGCCC TTTGTACACA
    P.pert_P  763  TCTTTGGTTT AAC-ATACCA AGAGGAATAC GTCCCTGCCC TTTGTACACA
  Allogrom_P  770  TCTT-GGTTC AAC-AAACCA CCAGGAATAT GTCCCTGCCC TTTGTACACA
       Align  388  mm----mmmm mmm-mmmmmm mmmmmmmmmm mmmmmmmmmm mmmmmmmmmm

     C_G.sI   921  CCGCCCGTCG CTCT-TACCG A--TG--AAT TGTA-CTG-T GAGTTT-GCA
     NA_G.sI  921  CCGCCCGTCG CTCT-TACCG A--TG--AAT TGTA-CTG-T GAGTTT-GCA
   C_G.sI(V)  921  CCGCCCGTCG CTCT-TACCG A--TG--AAT TGTA-CTG-T GAGTTT-GCA
    C_G.sIIa  902  CCGCCCGTCG CTCT-TACCG A--TG--AAT TTCC-CTG-T GAGTTT-GAA
   NA_G.sIIa  904  CCGCCCGTCG CTCT-TACCG A--TG--AAT TTCC-CTG-T GAGTTT-GAA
   CS_G.sIIa  891  CCGCCCGTCG CTCT-TACCG A--TG--AAT TTCC-CTG-T GAGTTT-GAA
   CA_G.sIIa  903  CCGCCCGTCG CTCT-TACCG A--TG--AAT TTCC-CTG-T GAGTTT-GAA
   CA_G.sIIb  893  CCGCCCGTCG CTCT-TACCG A--TG--ACT TTCC-CTG-T GAGTTT-GAA
   NA_G.sIIb  893  CCGCCCGTCG CTCT-TACCG A--TG--ACT TTCC-CTG-T GAGTTT-GAA
    NA_G.cal  913  CCGCCCGTCG CTCT-TACCG A--TG--AAT TTCA-CTG-T GAGTTC-AAT
     C_O.uni  861  CCGCCCGTCG CTCT-TACCG A-TTG--AAC TTTA-CTG-T GAGTTT-AAA
    CS_O.uni  861  CCGCCCGTCG CTCT-TACCG A-TTG--AAC TTTA-CTG-T GAGTTT-AAA
    CA_O.uni  843  CCGCCCGTCG CTCT-TACCG A-TTG--AAC TATA-CTG-C GAGTTT-GAA
  M_O.uni(V)  843  CCGCCCGTCG CTCT-TACCG A-TTG--AAC TATA-CTG-C GAGTTT-GAA
    C_G.sac   908  CCGCCCGTCG CTCT-TACCA A-TTG--AAC AGCA-CTA-C GAGTTT-AAA
   CS_G.sac   908  CCGCCCGTCG CTCT-TACCA A-TTG--AAC AGCA-CTA-C GAGTTT-AAA
  C_G.rub,P   887  CCGCCCGTCG CTCT-TACCG A--TG--GCT TTGG-CTG-C GAGTAA-AAG
 NA_G.rub,P   887  CCGCCCGTCG CTCT-TACCG A--TG--GCT TTGG-CTG-C GAGTAA-AAG
 NA_G.rub,WI  875  CCGCCCGTCG CTCT-TACCG A--TG--GCT GTGT-GTG-T GAGTAA-GAC
 CS_G.rub,WI  875  CCGCCCGTCG CTCT-TACCG A--TG--GCT GTGT-GTG-T GAGTAA-GAG
  C_G.rub(P)  880  CCGCCCGTCG CTCT-TACCG A--TG--GCT TTGT-GTG-T GAGTAA-GAC
NA_G.rub,WII  895  CCGCCCGTCG CTCT-TACCG A--TG--ACT TTGA-CTG-T GAGTAG-GAC
 C_G.rub,WII  895  CCGCCCGTCG CTCT-TACCG A--TG--ACT TTGA-CTG-T GAGTAG-GAC
  C_G.con(P)  923  CCGCCCGTCG CTCT-TACCG A--TG--ACT TTGA-CTG-T GAGTAA-GCT
   CS_G.con   923  CCGCCCGTCG CTCT-TACCG A--TG--ACT TTGA-CTG-T GAGTAG-GCT
```

```
 NA_N.pac(D)    998  CCGCCCGTCG CTCT-TACCG A--TG--AAC TCGT-TTG-T GAGTTT-TAA
 AA_N.pac(D)    998  CCGCCCGTCG CTCT-TACCG A--TG--AAC TCGT-TTG-T GAGTTT-TAA
     C_N.dut    901  CCGCCCGTCG CTCT-TACCG A--TG--AAC TTCT-TTG-T GAGTCT-AAG
 CS_G.bul,Ia    860  CCGCCCGTCG CTTT-TACCA A--TG--AAC TACT-TTG-C GAGAGT-GTG
CA_G.bul,IId    871  CCGCCCGTCG CTTT-TACCA A--TG--AAC TTCT-TTG-C GAGAGT-GAG
  M_G.bul(V)    862  CCGCCCGTCG CTTT-TACCA A--TG--AAC TACT-TTG-C GAGACT-ATG
 Na_G.bul,Ib    865  CCGCCCGTCG CTTT-TACCA A--TG--AAC TACT-TTG-C GAGACT-ATG
    NA_G.falc   878  CCGCCCGTCG CTTT-TACCG A--TG--AAC TTTT-TTG-C GAGTAT-GAT
    CS_G.falc   867  CCGCCCGTCG CTTT-TACCG A--TG--AAC TTTT-TTG-C GAGTAT-GAT
Na_G.bul,IIb    869  CCGCCCGTCG CTTT-TACCA A--TG--AAC TTCT-TTG-C GAGAGT-AAG
  AA_G.bul,IIb  869  CCGCCCGTCG CTTT-TACCA A--TG--AAC TTCT-TTG-C GAGAGT-AAG
Na-G.bul,IIa    864  CCGCCCGTCG CTTT-TACCA A--TG--AAC TTCT-TTG-C GAGAGT-GAG
  AA_G.bul,IIa  864  CCGCCCGTCG CTTT-TACCA A--TG--AAC TTCT-TTG-C GAGAGT-GAG
  AA_G.bul,IIc  867  CCGCCCGTCG CTTT-TACCA A--TG--AAC TTCT-TTG-C GAGAGT-GAG
NA-T.qui,IIa   1041  CCGCCCGTCG CTTT-TACCA A--TG--GCC CTCG-TTG-T GAGATA-GCT
 AA_T.qui,IIa  1041  CCGCCCGTCG CTTT-TACCA A--TG--GCC CTCG-TTG-T GAGATA-GCT
NA_T.qui,IIb   1054  CCGCCCGTCG CTTT-TACCA A--TG--GCC CTCG-TTG-T GAGATA-GCT
 AA_T.qui,IIb  1051  CCGCCCGTCG CTTT-TACCA A--TG--GCC CTCG-TTG-T GAGATA-GCT
  CS_T.qui,I    945  CCGCCCGTCG CTTT-TACCA A--TG--GCC CTCG-TTG-T GAGTGA-GCT
    CS_G.glu    846  CCGCCCGTCG CTCT-TACCG A--TG--GAC TTCT-CTG-T GAGTTT-GAG
    NA_G.uvu    898  CCGCCCGTCG CTCT-TACCG A--TG--GAC TTCT-CTG-T GAGTTT-GAG
   A.becca_P    884  CCGCCCGTCG CTCT-TACCG A--TG--GAT TATG-CTA-T GAATCT-ATA
     Troch_P    901  CCGCCCGTCG CTCT-TACCG A--TG--GAC TTCT-CTG-T GAGTTT-GAG
     H.germ_P   821  CCGCCCGTCG CTCT-TACCG A--TG--GAC TTCT-CTG-T GAGTTT-GAG
      Textu_P   879  CCGCCCGTCG CTCT-TACCG A--TG--GAC TTCT-CTG-T GAGTTT-GAG
      Biger_P   890  CCGCCCGTCG CTCT-TACCG A--TG--GAC TTCT-CTG-T GAGTTT-GAG
      Boliv_P   883  CCGCCCGTCG CTCT-TACCG A--TG--GAC TTCT-CTG-T GAGTTT-CAG
     G.oper_P   870  CCGCCCGTCG CTCT-TACCG A--TG--GAT TTCT-CTG-T GAGTTT-GAA
    A.trian_P  1261  CCGCCCGTCG CTCT-TACCG A--TG--GAC TACT-CTG-T GAGTTT-GAG
    As.rara_P  1255  CCGCCCGTCG CTCT-TACCG A--TG--GAC TACT-CTG-T GAGTTT-AAG
     A.angu_P   690  CCGCCCGTCG CTCT-TACCG A--TG--AAT TATA-TTA-T AAATCT-AAG
     E.acul_P   649  CCGCCCGTCG CTCT-TACCG A--TG--AAC TTCG-CTA-T GAATCT-ATT
       Quin_P   815  CCGCCCGTCG CTCT-TACCG A--TG--AAT TATA-TTA-T AAATCT-AAG
    M.secan_P   795  CCGCCCGTCG CTCT-TACCG A--TG--AAT TATA-TTA-T AAATCT-AAG
     P.pert_P   812  CCGCCCGTCG CTCT-TACCG A--TG--AAT TATA-TTA-T AAATCT-AAG
   Allogrom_P   818  CCGCCCGTCG CTCT-TACCG A--TG--GAC TATT-CTG-T GAGTAT-ACA
         Align  433  mmmmmmmmmm mmmm-mmmm m--mm--mmm mmmm-mmm-m mmmmmm-mmm

       C_G.sI   963  GGACCGA--- --ACCCA--- ---------- ---------- ----------
      NA_G.sI   963  GGACCGA--- --ACCCA--- ---------- ---------- ----------
    C_G.sI(V)   963  GGACCGA--- --ACCCA--- ---------- ---------- ----------
     C_G.sIIa   944  GGACTGA--- --TGGTTG-A AAATC----- ---------- ----------
    NA_G.sIIa   946  GGACTGA--- --TGGTTGCA AAATC----- ---------- ----------
    CS_G.sIIa   933  GGACTGA--- --TGGTTGNA AAATC----- ---------- ----------
    CA_G.sIIa   945  GGACTGA--- --TGGTTGCA AAATC----- ---------- ----------
    CA_G.sIIb   935  GGACTGG--- --TGGTTGCA TGACT----- ---------- ----------
    NA_G.sIIb   935  GGACTGG--- --TGGTTGCA TGACT----- ---------- ----------
     NA_G.cal   955  GGACCGA--- --TTTTTTCC C--------- ---------- ----------
      C_O.uni   904  GGACCGA--- --TCA----- ---------- ---------- ----------
     CS_O.uni   904  GGACCGA--- --TCA----- ---------- ---------- ----------
     CA_O.uni   886  GGACCAG--- --GTCTCTGG CA-------- ---------- ----------
   M_O.uni(V)   886  GGACCAG--- --GTCTCTGG CA-------- ---------- ----------
      C_G.sac   951  GGCCCGA--- --GAA----- ---------- ---------- ----------
     CS_G.sac   951  GGCCCGA--- --GAA----- ---------- ---------- ----------
    C_G.rub,P   929  GGACTTT--- --CG------ ---------- ---------- ----------
   NA_G.rub,P   929  GGACTTT--- --CG------ ---------- ---------- ----------
  NA_G.rub,WI   917  AGACTGT--- --TA------ ---------- ---------- ----------
  CS_G.rub,WI   917  CGACTGT--- --TA------ ---------- ---------- ----------
    C_G.rub(P)  922  GGACAAT--- --TA------ ---------- ---------- ----------
 NA_G.rub,WII   937  TGACCGT--- --TTA----- ---------- ---------- ----------
  C_G.rub,WII   937  TGACCGT--- --TTA----- ---------- ---------- ----------
    C_G.con(P)  965  GGATCGA--- ---------- ---------- ---------- ----------
     CS_G.con   965  GGATCGA--- ---------- ---------- ---------- ----------
  NA_N.pac(D)  1040  GGACTGG--- --ATTAAGCT ATATGCGTAA C--------- ----------
  AA_N.pac(D)  1040  GGACTGG--- --ATTAAGCT ATATGCGTAA C--------- ----------
      C_N.dut   943  GGACTGG--- --GTTAATCT GTAATTCTAT TATAGATATA CC--------
  CS_G.bul,Ia   902  GGACCTA--- --TGAGCTTT CAAG------ ---------- ----------
 CA_G.bul,IId   913  AGACTTA--- --AAGATA-- ---------- ---------- ----------
   M_G.bul(V)   904  G-ACTT-GA- --AGCATAGG ATTGAACTGC ---------- ----------
  Na_G.bul,Ib   907  GGACTTTGA- --AGCATAGG ATTTGAACTG GC-------- ----------
    NA_G.falc   920  GGACTAA--- --CTTTGCGC TCAATTTGAA ---------- ----------
    CS_G.falc   909  GGACTAA--- --CTTTGCGC TCAATTTGAA ---------- ----------
 Na_G.bul,IIb   911  AGACTTG--- --TAAT---- ---------- ---------- ----------
  AA_G.bul,IIb  911  AGACTTG--- --TAAT---- ---------- ---------- ----------
Na-G.bul,IIa    906  AGACTAA--- --AGAT---- ---------- ---------- ----------
  AA_G.bul,IIa  906  AGACTAA--- --AGAT---- ---------- ---------- ----------
  AA_G.bul,IIc  909  AGACTGA--- --AATGT--- ---------- ---------- ----------
```

```
NA-T.qui,IIa  1083  GGACAAG--- --TATTTA-- ---------- ---------- ----------
 AA_T.qui,IIa 1083  GGACAAG--- --TATTTA-- ---------- ---------- ----------
NA_T.qui,IIb  1096  GGACAAG--- --TATTTA-- ---------- ---------- ----------
 AA_T.qui,IIb 1093  GGACAAG--- --TATTTA-- ---------- ---------- ----------
  CS_T.qui,I   987  GGACAAG--- --TTATTT-- ---------- ---------- ----------
    CS_G.glu   888  GGACTGG--- --ATTTGATC -GCTTCGGCG CATTAATAT- ----------
    NA_G.uvu   940  GGACTGG--- --AGCTTACT TTACTGTGAG TCT------- ----------
   A.becca_P   926  GGACTGC--- --CAAAGTTG CCTCGGCTGC TT-------- ----------
     Troch_P   943  GGACTGG--- --ATAT---T GTAATTTCGG TTACTACCAT C---------
    H.germ_P   863  GGACTGG--- --ATTTTATA TC-------- ---------- ----------
     Textu_P   921  GGACTGG--- --GTACTGCT ATAAATTTAT TTACTGCTAT CACC------
     Biger_P   932  GGACTGG--- --GTACTGTA AAAATTTATT TTTACGATCA CC--------
     Boliv_P   925  GGACTGT--- --CTTTGGCG CTTCATATAT TACGCTTCGG CGCTATATAT
    G.oper_P   912  GGACTGG--- --CCTTCTGT GC-------- ---------- ----------
   A.trian_P  1303  GGACTGG--- --TTTGAATT TTAAAGTATA TGTATATTTA TTTATATATA
   As.rara_P  1297  GGACTGG--- --TTTGAATT AATATAATTT TTTTAGAAAA TTTATTTTTT
    A.angu_P   732  GGACAAT--- --AATTAATA TAAT------ ---------- ----------
    E.acul_P   691  AGACTGC--- --GTTATACG ---------- ---------- ----------
      Quin_P   857  GGACTTA--- --AAATATAT TTTATATAT- ---------- ----------
   M.secan_P   837  GGACTTA--- --AAATATAA TTTATTATAT ---------- ----------
    P.pert_P   854  GGACATA--- --CATATTCT GAAATTGAAT A--------- ----------
  Allogrom_P   860  GGACGGG--- --AACTCTAT CTTCTGATTG AGTTC----- ----------
       Align   475  mmmmmmm--- ---------- ---------- ---------- ----------

        C_G.sI   975  ---------- ---------- ---------- ---------- ----TTTTTG
       NA_G.sI   975  ---------- ---------- ---------- ---------- ----TTTTTG
     C_G.sI(V)   975  ---------- ---------- ---------- ---------- ----TTTTTG
      C_G.sIIa   963  ---------- ---------- ---------- ---------- ----------
     NA_G.sIIa   966  ---------- ---------- ---------- ---------- ----------
     CS_G.sIIa   953  ---------- ---------- ---------- ---------- ----------
     CA_G.sIIa   965  ---------- ---------- ---------- ---------- ----------
     CA_G.sIIb   955  ---------- ---------- ---------- ---------- ----------
     NA_G.sIIb   955  ---------- ---------- ---------- ---------- ----------
      NA_G.cal   971  ---------- ---------- ---------- ---------- ----------
        C_O.uni  914  ---------- ---------- ---------- ---------- ----------
       CS_O.uni  914  ---------- ---------- ---------- ---------- ----------
       CA_O.uni  903  ---------- ---------- ---------- ---------- ----------
    M_O.uni(V)   903  ---------- ---------- ---------- ---------- ----------
        C_G.sac  961  ---------- ---------- ---------- ---------- ----------
       CS_G.sac  961  ---------- ---------- ---------- ---------- ----------
     C_G.rub,P   938  ---------- ---------- ---------- ---------- ----------
    NA_G.rub,P   938  ---------- ---------- ---------- ---------- ----------
   NA_G.rub,WI   926  ---------- ---------- ---------- ---------- ----------
   CS_G.rub,WI   926  ---------- ---------- ---------- ---------- ----------
    C_G.rub(P)   931  ---------- ---------- ---------- ---------- ----------
  NA_G.rub,WII   947  ---------- ---------- ---------- ---------- ----------
   C_G.rub,WII   947  ---------- ---------- ---------- ---------- ----------
    C_G.con(P)   972  ---------- ---------- ---------- ---------- ----------
       CS_G.con  972  ---------- ---------- ---------- ---------- ----------
   NA_N.pac(D)  1066  ---------- ---------- ---------- ---------- ----------
   AA_N.pac(D)  1066  ---------- ---------- ---------- ---------- ----------
        C_N.dut  980  ---------- ---------- ---------- ---------- ----------
   CS_G.bul,Ia   921  ---------- ---------- ---------- ---------- ----------
  CA_G.bul,IId   926  ---------- ---------- ---------- ---------- ----------
    M_G.bul(V)   929  ---------- ---------- ---------- ---------- ----------
   Na_G.bul,Ib   936  ---------- ---------- ---------- ---------- ----------
      NA_G.falc  945  ---------- ---------- ---------- ---------- ---------T
      CS_G.falc  934  ---------- ---------- ---------- ---------- ---------T
  Na_G.bul,IIb   922  ---------- ---------- ---------- ---------- ----------
   AA_G.bul,IIb 922  ---------- ---------- ---------- ---------- ----------
  Na-G.bul,IIa   917  ---------- ---------- ---------- ---------- ----------
    AA_G.bul,IIa 917  ---------- ---------- ---------- ---------- ----------
    AA_G.bul,IIc 921  ---------- ---------- ---------- ---------- ---------C
NA-T.qui,IIa  1096  ---------- ---------- ---------- ---------- ---------C
 AA_T.qui,IIa 1096  ---------- ---------- ---------- ---------- ---------C
NA_T.qui,IIb  1109  ---------- ---------- ---------- ---------- ---------C
 AA_T.qui,IIb 1106  ---------- ---------- ---------- ---------- ---------C
  CS_T.qui,I  1000  ---------- ---------- ---------- ---------- ---------T
    CS_G.glu   921  ---------- ---------- ---------- ---------- ----------
    NA_G.uvu   968  ---------- ---------- ---------- ---------- ----------
   A.becca_P   953  ---------- ---------- ---------- ---------- ----------
     Troch_P   976  ---------- ---------- ---------- ---------- ----------
    H.germ_P   880  ---------- ---------- ---------- ---------- ----------
     Textu_P   960  ---------- ---------- ---------- ---------- ----------
     Biger_P   969  ---------- ---------- ---------- ---------- ----------
     Boliv_P   970  ACGCCATC-- ---------- ---------- ---------- ----------
    G.oper_P   929  ---------- ---------- ---------- ---------- ----------
```

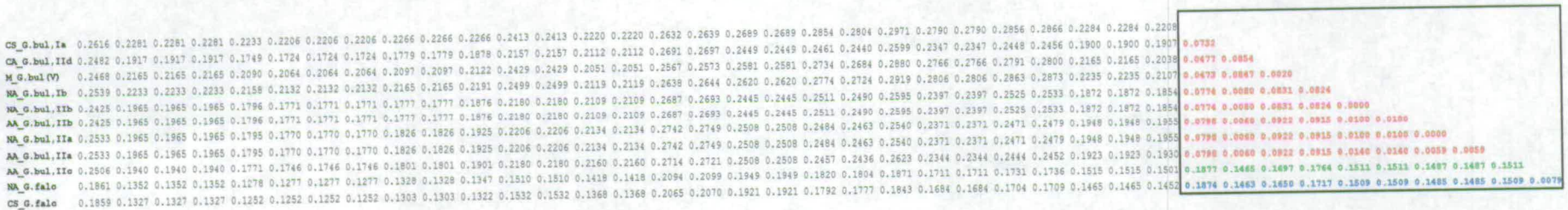
```
A.trian_P  1348  TTACTTTTTA TATAATTCAG GC-------- ---------- ----------
As.rara_P  1342  ATTATTATAT TATATTATAT TCAGGC---- ---------- ----------
 A.angu_P   751  ---------- ---------- ---------- ---------- ----------
 E.acul_P   706  ---------- ---------- ---------- ---------- ----------
   Quin_P   881  ---------- ---------- ---------- ---------- ----------
M.secan_P   862  ---------- ---------- ---------- ---------- ----------
 P.pert_P   880  ---------- ---------- ---------- ---------- ----------
Allogrom_P  890  ---------- ---------- ---------- ---------- ----------
    Align   482  ---------- ---------- ---------- ---------- ----------

   C_G.sI   981  GG-TTTGGAA AT---GCAG- TCAAA-CAGT A-CGATTTAA -AGGAAAG-A
  NA_G.sI   981  GG-TTTGGAA AT---GCAG- TCAAA-CAGT A-CGATTTAA -AGGAAAG-A
C_G.sI(V)   981  GG-TTTGGAA AT---GCAG- TCAAA-CAGT A-CGATTTAA -AGGAAAG-A
 C_G.sIIa   963  ---ATTGGAA AT---TCTG- TCAAA-CAGC G-AGATTTAA -AGGAAAG-A
NA_G.sIIa   966  ---ATTGGAA AT---TCTG- TCAAA-CAGC G-AGATTTAA -AGGAAAG-A
CS_G.sIIa   953  ---ATTGGAA AT---TCTG- TCAAA-CAGC G-AGATTTAA -AGGAAAG-A
CA_G.sIIa   965  ---ATTGGAA AT---TCTG- TCAAA-CAGC G-AGATTTAA -AGGAAAG-A
CA_G.sIIb   955  ---ATTGGAA AT---TCTG- TCAAA-CAGC G-AGATTTAA -AGGAAAG-A
NA_G.sIIb   955  ---ATTGGAA AT---TCTG- TCAAA-CAGC G-AGATTTAA -AGGAAAG-A
 NA_G.cal   971  ---TTTGGAA AT---TTGG- TCAAA-CAGT G-AGATTTAA -AGGAAAG-A
   C_O.uni  914  ---TTTGGAA AT---TTAG- TCAAA-CAGA G-TTGTTTAA -AGGAAAG-A
  CS_O.uni  914  ---TTTGGAA AT---TTAG- TCAAA-CAGA G-TTGTTTAA -AGGAAAG-A
  CA_O.uni  903  ---CTTGGAA AT---TTAG- TCAAA-CAGA G-TTGTTTAA -AGGAAAG-A
M_O.uni(V)  903  ---CTTGGAA AT---TTAG- TCAAA-CAGA G-TTGTTTAA -AGGAAAG-A
   C_G.sac  961  ---ATTGGAA AT---TTAG- TCAAA-CAGT G-CTGTTTAA -AGGAAAG-A
  CS_G.sac  961  ---ATTGGAA AT---TTAG- TCAAA-CAGT G-CTGTTTAA -AGGANAG-A
 C_G.rub,P  938  --TTTTGGAA CT----TTG- TCGAA-CAGA T-GGGGCTAA -AGGAAAG-A
NA_G.rub,P  938  --TTTTGGAA CT----TTG- TCGAA-CAGA T-GGGGCTAA -AGGAAAG-A
NA_G.rub,WI 926  --AGAGGGAA AT----TTG- TCGAA-TGCA T-TTGGCTAA -AGGAAAG-A
CS_G.rub,WI 926  --AGAGGGAA AT----TTG- TCGAA-TGCA T-TTGGCTAA -AGGAAAG-A
 C_G.rub(P) 931  --AGCTGGAA AT----TTG- TCGAA-TGCA T-TGGGCTAA -AGGAAAG-A
NA_G.rub,WII 947 ---TCGGGAA AT----CCG- TCGAA-CAGT T-AGAGTTAA -AGGAAAG-A
C_G.rub,WII 947  ---TCGGGAA AT----CCG- TCGAA-CAGT T-AGAGTTAA -AGGAAAG-A
 C_G.con(P) 972  --AATGGGAA AG----CTG- TCAAA-CAGT T-AGAGTTAA -AGGAAAG-A
   CS_G.con 972  --AATGGGAA AG----CTG- TCAAA-CAGT T-AGAGTTAA -AGGAAAG-A
NA_N.pac(D) 1066 --CTATGGAA AT---TCAT- GCGA--TGAA C-TGGTTTAA -AGGAAAG-A
AA_N.pac(D) 1066 --CTATGGAA AT---TCAT- GCGA--TGAA C-TGGTTTAA -AGGAAAG-A
   C_N.dut  980  ---TATGGAA AC---TTAT- ACGAA-CAAT G-TGGTTTAA -AGGAAAG-A
CS_G.bul,Ia 921  -TACAGGGAA CC----CAT- TCGAC-CAAC G-GAGTTTAA -AGGAAAA-A
CA_G.bul,IId 926 -ATCGCGG-A AC---TCAC- TCGAC-CGAC G-GGATTTAA -AGGAAAA-A
 M_G.bul(V) 929  ---GATGGAA AT---GTAT- TCGAT-CAAC G-GAGTTTAA -AGGAAAA-A
Na_G.bul,Ib 936  ---GATGGGA AAT--GTAT- TCGAT-CAAC G-GAGTTTAA -AGGAAAA-A
  NA_G.falc 946  CTGTTGTGGA AA--TTCAC- TCGAA-CAAC G-GGATTTAA -AGGAAAA-A
  CS_G.falc 935  CTGTTGTGGA AA--TTCAC- TCGAA-CAAC G-GGATTTAA -AGGAAAA-A
Na_G.bul,IIb 922 AATCGTGG-A AC---TCAC- TCGAC-CGAC G-GGATTTAA -AGGAAAA-A
AA_G.bul,IIb 922 AATCGTGG-A AC---TCAC- TCGAC-CGAC G-GGATTTAA -AGGAAAA-A
Na-G.bul,IIa 917 -ATCGCGG-A AC---TCAC- TCGAC-CGAC G-GGACTTAA -AGGAAAA-A
AA_G.bul,IIa 917 -ATCGCGG-A AC---TCAC- TCGAC-CGAC G-GGACTTAA -AGGAAAA-A
AA_G.bul,IIc 921 --GTGTGG-A AC---TCAC- TCGAC-CGAC G-GGACTTAA -AGGAAAA-A
NA-T.qui,IIa 1097 TACTTCAAAA GT---TC--- TCAAT-CAAG G-TTTGCTAA -AGGAAAA-A
AA_T.qui,IIa 1097 TACTTCAAAA GT---TC--- TCAAT-CAAG G-TTTGCTAA -AGGAAAA-A
NA_T.qui,IIb 1110 TACTTCAAAA GT---TC--- TCAAT-CAAG G-TTTGCTAA -AGGAAAA-A
AA_T.qui,IIb 1107 TACTTCAAAA GT---TC--- TCAAT-CAAG G-TTTGCTAA -AGGAAAA-A
 CS_T.qui,I 1001 AACTTGAAAA GT---TC--- TCAAT-CAAG G-TTTGCTAA -AGGAAAA-A
   CS_G.glu  921  --CTATGGAA AC---TTAA- ACGAA-CAGT G-TGGTCTAA -AGGAAAG-A
   NA_G.uvu  968  --CTATGGAA AC---TTAA- ACGAA-CAGT G-TGGTCTAA -AGGAAAG-A
  A.becca_P  953  ---AGTGGAA AT---ATAT- ATGAA-TAGC G-TGATCTAA -AGGAAAG-A
    Troch_P  976  ---TATGGAA AC---TTAA- ACGAA-CAGT G-TGGTCTAA -AGGAAAG-A
   H.germ_P  880  ---TATGGAA AC---TCAA- ACGAA-CAGT G-TGGTCTAA -AGGAAAG-A
    Textu_P  960  ---TACGGAA AC---TTAA- ACGAA-CAGT G-TGGTCTAA -AGGAAAG-A
    Biger_P  969  ---TATGGAA AC---TTAA- ACGAA-CAGT G-TGGTCTAA -AGGAAAG-A
    Boliv_P  978  ---TTAGGAA AC---TTAA- ACGAA-CAGT G-TGGTCTAA -AGGAAAG-A
   G.oper_P  929  ---TATGGAA AN---TCAA- ACGAA-CAGT G-TGATCTAA -AGGAAAG-A
  A.trian_P 1370  ---TATGGAA AC---TCAT- ACGAA-CAGT G-TGGTCTAA -AGGAAAG-A
  As.rara_P 1368  ---TATGGAA AC---TTAT- ACGAA-CAGT G-TGGTCTAA -AGGAAAG-A
   A.angu_P  751  ---TTATGAA AC---ATAT- ATACA-TAAT G-TGATTTAA -AGGAAAG-A
   E.acul_P  706  ---CGCGGAA AG---ATAT- ATGAA-TAGT G-TGGTTTAA -AGGAAAG-A
     Quin_P  881  ---TTCGGAA AC---TTAT- ATGCA-TAAT G-TGATTTAA -AGGAAAG-A
  M.secan_P  862  ---TTCGGAA AC---TTAT- ATGCA-TAAT G-TGATTTAA -AGGAAAG-A
   P.pert_P  880  ---TATTGAA AC---TTAT- ATACA-TAAT G-TGATTTAA -AGGAAAG-A
 Allogrom_P  890  ---TATAAGA AT---GTAC- GCGAA-CAGT G-TGGTCTAA -AGGAAAG-A
      Align  482  ------mm-m mm-------- mmmm--mmmm m-mmmmmmmm -mmmmmmm-m

                                    C_G.sI  1022  GAA-G-TCGT AAC
                                   NA_G.sI  1022  GAA-G-TCGT AAC
                                 C_G.sI(V)  1022  GAA-G-TCGT AAC
                                  C_G.sIIa  1002  GAA-G-TCGT AAC
```

```
NA_G.sIIa  1005  GAA-G-TCGT AAC
CS_G.sIIa   992  GAA-G-TCGT AAC
CA_G.sIIa  1004  GAA-G-TCGT AAC
CA_G.sIIb   994  GAA-G-TCGT AAC
NA_G.sIIb   994  GAA-G-TCGT AAC
NA_G.cal   1010  GAA-G-TCGT AAC
C_O.uni     953  GAA-G-TCGT AAC
CS_O.uni    953  GAA-G-TCGT AAC
CA_O.uni    942  GAA-G-TCGT AAC
M_O.uni(V)  942  GAA-G-TCGT AAC
C_G.sac    1000  GAA-G-TCGT AAC
CS_G.sac   1000  GAA-G-TCGT AAC
C_G.rub,P   977  GAA-G-TCGT AAC
NA_G.rub,P  977  GAA-G-TCGT AAC
NA_G.rub,WI 965  GAA-G-TCGT AAC
CS_G.rub,WI 965  GAA-G-TCGT AAC
C_G.rub(P)  970  GAA-G-TCGT AAC
NA_G.rub,WII 985 GAA-G-TCGT AAC
C_G.rub,WII 985  GAA-G-TCGT AAC
C_G.con(P) 1011  GAA-G-TCGT AAC
CS_G.con   1011  GAA-G-TCGT AAC
NA_N.pac(D) 1105 GAA-G-TCGT AAC
AA_N.pac(D) 1105 GAA-G-TCGT AAC
C_N.dut    1019  GAA-G-TCGT AAC
CS_G.bul,Ia  961 GAA-G-TCGT AAC
CA_G.bul,IId 966 GAA-G-TCGT AAC
M_G.bul(V)   968 GAA-G-TCGT AAC
Na_G.bul,Ib  976 GAA-G-TCGT AAC
NA_G.falc    989 GAA-G-TCGT AAC
CS_G.falc    978 GAA-G-TCGT AAC
Na_G.bul,IIb 963 GAA-G-TCGT AAC
AA_G.bul,IIb 963 GAA-G-TCGT AAC
Na-G.bul,IIa 957 GAA-G-TCGT AAC
AA_G.bul,IIa 957 GAA-G-TCGT AAC
AA_G.bul,IIb 960 GAA-G-TCGT AAC
NA-T.qui,IIa 1137 GAA-G-TCGT AAC
AA_T.qui,IIa 1137 GAA-G-TCGT AAC
NA_T.qui,IIb 1150 GAA-G-TCGT AAC
AA_T.qui,IIb 1147 GAA-G-TCGT AAC
CS_T.qui,I  1041  GAA-G-TCGT AAC
CS_G.glu     961  GAA-G-TCGT AAC
NA_G.uvu    1008  GAA-G-TCGT AAC
A.becca_P    992  GAA-G-TCGT AAC
Troch_P     1015  GAA-G-TCGT AAC
H.germ_P     919  GAA-G-TCGT AAC
Textu_P      999  GAA-G-TCGT AAC
Biger_P     1008  GAA-G-TCGT AAC
Boliv_P     1017  GAA-G-TCGT AAC
G.oper_P     968  GAA-G-TCGT AAC
A.trian_P   1409  GAA-G-TCGT AAC
As.rara_P   1407  GAA-G-TCGT AAC
A.angu_P     790  GAA-G-TCGT AAC
E.acul_P     745  GAA-G-TCGT AAC
Quin_P       920  GAA-G-TCGT AAC
M.secan_P    901  GAA-G-TCGT AAC
P.pert_P     919  GAA-G-TCGT AAC
Allogrom_P   929  GAA-G-TCGT AAC
Align        512  mmm-m-mmmm mmm
```

A1.3. Distance matrix for 505 bp molecular phylogeny

| | | |

A1.4. Calculating evolutionary distance from the distance matrix



	CS- Ib	CA- IId	M- (V)	NA- Ib	NA- IIB	AA-IIB	NA- IIA	AA- IIA	AA- IIC	NA-falc
CS_G.bul,Ia										
CA_G.bul,IId	0.0732									
M_G.bul(V)	0.0477	0.0854								
NA_G.bul,Ib	0.0473	0.0847	0.0020							
NA_G.bul,IIb	0.0774	0.0080	0.0831	0.0824						
AA_G.bul,IIb	0.0774	0.0080	0.0831	0.0824	0.0000					
NA_G.bul,IIa	0.0798	0.0060	0.0922	0.0915	0.0100	0.0100				
AA_G.bul,IIa	0.0798	0.0060	0.0922	0.0915	0.0100	0.0100	0.0000			
AA_G.bul,IIc	0.0798	0.0060	0.0922	0.0915	0.0140	0.0140	0.0059	0.0059		
NA_G.falc	0.1877	0.1465	0.1697	0.1764	0.1511	0.1511	0.1487	0.1487	0.1511	
CS_G.falc	0.1874	0.1463	0.1650	0.1717	0.1509	0.1509	0.1485	0.1485	0.1509	0.0079

This is a section of the distance matrix shown in A1.3. The boxed area is expanded for easier viewing and represents the evolutionary distances between the *Globigerina* sp. genotypes. The evolutionary distances are calculated out of a total of one. The method by which the mean evolutionary distances are calculated between the *G. bulloides* and the *G. falconensis* monophyletic groups is shown overleaf.
Abbreviations: the nucleotide sequences are denoted as in the alignment (A1.1).

Evolutionary distance calculation:

1. The meanevolutionary distance between the North Atlantic *G. falconensis* genotype and the *G. bulloides* monophyletic group is represented by the green numbers on the matrix. Therefore,

NA_*G. falc* vs. *G. bulloides* = (0.1877 + 0.1465 + 0.1697 + 0.1764 + 0.1511 + 0.1511 + 0.1487 + 0.1487 + 0.1511) / 9

= 1.431 / 9

= 0.159 (out of a total of 1)

2. The mean evolutionary distance between the Coral Sea *G. falconensis* genotype and the *G. bulloides* monophyletic group is represented by the blue numbers on the matrix. Therefore,

CS_*G. falc* vs. *G. bulloides* = (0.1874 + 0.1463 + 0.1650 + 0.1717 + 0.1509 + 0.1509 + 0.1485 + 0.1485 + 0.1509) / 9

= 1.420 / 9

= 0.1578 (out of a total of 1)

3. The mean evolutionary distance between the *G. falconensis* and the *G. bulloides* monophyletic groups is therefore,

= (0.159 + 0.1578) / 2

= 0.1584 (out of a total of one)

= **15.8 %**

Appendix 2.1 Within morphospecies nucleotide sequence alignments

A2.1.1. *Globigerina bulloides* 851 bp molecular phylogeny

```
CA-IId      1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
NA-IIb      1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
AA-IIb      1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
AA-IIc      1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
AA-IIa      1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
NA-IIa      1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
  CS-I      1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
 Align      1  mmm--mmmmm mm--mmmmm- --mmmmmmm-m mmmmmmmmmmm mmmmmmmmmmm

CA-IId     43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACAC-AGTGA G--GATTG-A
NA-IIb     43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACGC-AGTGA G--GATTG-A
AA-IIb     43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACGC-AGTGA G--GATTG-A
AA-IIc     43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACAC-AGTGA G--GATTG-A
AA-IIa     43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACGC-AGTGA G--GATTG-A
NA-IIa     43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACGC-AGTGA G--GATTG-A
  CS-I     43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACAC-AGTGA G--GATTG-A
 Align     43  mmmmm-mmmm mmmmmmmmm- -----mmmmm mmmm-mmmmm m--mmmmm-m

CA-IId     82  CAGACAGTT- ------AGAC AGAAGTGAGG TTTT-GGTAA ---CAA----
NA-IIb     82  CAGACAGTT- --------AG ACAGAAGTGG TTT-AGGTAA ---CAA----
AA-IIb     82  CAGACAGTT- --------AG ACAGAAGTGG TTT-AGGTAA ---CAA----
AA-IIc     82  CAGACAGTT- ---------G GCAGGAGTGG TTCTTGGTAA AAACAA----
AA-IIa     82  CAGACAGTT- ---------- TCAGGAGTGG TTCTTGGTAA A--CAA----
NA-IIa     82  CAGACAGTT- ---------- TCAGGAGTGG TTCTTGGTAA A--CAA----
  CS-I     82  CAGACGGTC- ---------- TTATTGGTGG ACTCTAAGAC GTC-------
 Align     82  mmmmmmmmm- ---------- --------mm m--------- ----------

CA-IId    117  ---------- ---------- ---------- ---------- ---------T
NA-IIb    115  ---------- ---------- ---------- ---------- ---------T
AA-IIb    115  ---------- ---------- ---------- ---------- ---------T
AA-IIc    118  ---------- ---------- ---------- ---------- ---------A
AA-IIa    115  ---------- ---------- ---------- ---------- ---------T
NA-IIa    115  ---------- ---------- ---------- ---------- ---------T
  CS-I    114  ---------- ---------- ---------- ---------- ---------A
 Align     94  ---------- ---------- ---------- ---------- ----------

CA-IId    118  TGAGAGAGTT GAA-GT---- ---------- -----TCTTT CATGA-----
NA-IIb    116  TGAGAGAGTT GAA-GT---- ---------- -----TCTTT CATGA-----
AA-IIb    116  TGAGAGAGTT GAA-GT---- ---------- -----TCTTT CATGA-----
AA-IIc    119  TGAGAGAGTT GAA-GT---- ---------- -----TCTTT CATGA-----
AA-IIa    116  TGAGAGAGTT GAA-GT---- ---------- -----TCTTT CATGA-----
NA-IIa    116  TGAGAGAGTT GAA-GT---- ---------- -----TCTTT CATGA-----
  CS-I    115  TGGTAAAGTT GAA-GT---- ---------- -----TCTTT CATGA-----
 Align     94  mmmmmmmmmmm mmm-mm---- ---------- -----mmmmm mmmmm-----

CA-IId    143  -TCTTGTGAG AG-----GTG GTG-CATGG- CCGTT-CTTA GTTCGTGTA-
NA-IIb    141  -TCTTGTGAG AG-----GTG GTG-CATGG- CCGTT-CTTA GTTCGTGTA-
AA-IIb    141  -TCTTGTGAG AG-----GTG GTG-CATGG- CCGTT-CTTA GTTCGTGTA-
AA-IIc    144  -TCTTGTGAG AG-----GTG GTG-CATGG- CCGTT-CTTA GTTCGTGTA-
AA-IIa    141  -TCTTGTGAG AG-----GTG GTG-CATGG- CCGTT-CTTA GTTCGTGTA-
NA-IIa    141  -TCTTGTGAG AG-----GTG GTG-CATGG- CCGTT-CTTA GTTCGTGTA-
  CS-I    140  -TCTTGTGGC AG-----GTG GTG-CATGG- CCGTT-CTTA GTTCGTGTA-
 Align    119  -mmmmmmmmm mm-----mmm mmm-mmmmm- mmmmm-mmmm mmmmmmmmm-

CA-IId    183  GTGATA-TGT C--TGCCT-A ATCGCGTCAC -------GAT AACCTATTGG
NA-IIb    181  GTGATA-TGT C--TGCCT-A ATCGCGTCAC -------GAT AACCTATTGG
AA-IIb    181  GTGATA-TGT C--TGCCT-A ATCGCGTCAC -------GAT AACCTATTGG
AA-IIc    184  GTGATA-TGT C--TGCCT-A ATCGCGTCAC -------GAT AACCTATTGG
AA-IIa    181  GTGATA-TGT C--TGCCT-A ATCGCGTCAC -------GAT AACCTATTGG
NA-IIa    181  GTGATA-TGT C--TGCCT-A ATCGCGTCAC -------GAT AACCTATTGG
  CS-I    180  GTGATA-CGT C--TGCCT-A ATCGCGTCAC -------GAT AACCTATTTC
 Align    159  mmmmmm-mmm m--mmmmm-m mmmmmmmmmmm -------mmm mmmmmmmmmm

CA-IId    222  TCGACAACCC AATTATC-AC AACTGCAGCA TAACTCCCC- TTGGGTGGGC
NA-IIb    220  TCGACAACCC AATTATC-AC AACTGCAGCA TAACTCCCC- TTGGGTGGGC
AA-IIb    220  TCGACAACCC AATTATC-AC AACTGCAGCA TAACTCCCC- TTGGGTGGGC
AA-IIc    223  TCGACAACCC AATTATC-AC AACTGCAGCA TAACTCCCC- TTGGGTGGGC
```

```
AA-IIa   220  TCGACAACCC AATTATC-AC AACTGCAGCA TAACTCCCC- TTGGGTGGGC
NA-IIa   220  TCGACAACCC AATTATC-AC AACTGCAGCA TAACTCCCC- TTGGGTGGGC
  CS-I   219  TCGACAATCC AAAGATC-TT CG-TGCAGTA T--CTA---- TTGGTACGGT
 Align   198  mmmmmmmmmmmm mmmmmmmm--- ---mmmmmmm m--------- mmmmmmmmmmm

CA-IId   270  GAGGCTCTGT GTAGGATAGA C--------- ---------- ----------
NA-IIb   268  GAGGCTCTGT GTAGGATAGA C--------- ---------- ----------
AA-IIb   268  GAGGCTCTGT GTAGGATAGA C--------- ---------- ----------
AA-IIc   271  GAGGCTCTGT GTAGGATAGA C--------- ---------- ----------
AA-IIa   268  GAGGCTCTGT GTAGGATAGA C--------- ---------- ----------
NA-IIa   268  GAGGCTCTGT GTAGGATAGA C--------- ---------- ----------
  CS-I   261  GCTCACGTTA GTAGGCTAGA G--------- ---------- ----------
 Align   233  mmmmmmmmmmm mmmmmmmmmmm m--------- ---------- ----------

CA-IId   291  ---------- ---------- ---------- ---------- -----CTCTG
NA-IIb   289  ---------- ---------- ---------- ---------- -----CTCTG
AA-IIb   289  ---------- ---------- ---------- ---------- -----CTCTG
AA-IIc   292  ---------- ---------- ---------- ---------- -----CTCTG
AA-IIa   289  ---------- ---------- ---------- ---------- -----CTCTG
NA-IIa   289  ---------- ---------- ---------- ---------- -----CTCTG
  CS-I   282  ---------- ---------- ---------- ---------- -----ATTTG
 Align   254  ---------- ---------- ---------- ---------- -----mmmmm

CA-IId   296  AACAGTACGC AACGAACGCG ATCGT-AA-- --TCCCTTGT --TG------
NA-IIb   294  AACAGTACGC AACGAACGCG ATCGT-AA-- --TCCCTTGT --TG------
AA-IIb   294  AACAGTACGC AACGAACGCG ATCGT-AA-- --TCCCTTGT --TG------
AA-IIc   297  AACAGTACGC AACGAACGCG ATCGT-AA-- --TCCCTTGT --TG------
AA-IIa   294  AACAGTACGC AACGAACGCG ATCGT-AA-- --TCCCTTGT --TG------
NA-IIa   294  AACAGTACGC AACGAACGCG ATCGT-AA-- --TCCCTTGT --TG------
  CS-I   287  AACAGTACGC AACGGACGCG ATCGT-AA-- --TCTCTTGT --TA------
 Align   259  mmmmmmmmmmm mmmmmmmmmmm mmmmm-mm-- --mmmmmmmm --mm------

CA-IId   333  ---------- -AGTGGCCAT CCTGTAAGC- -TGCTGGATT AGGAACC---
NA-IIb   331  ---------- -AGTGGCCAT CCTGTAAGC- -TGCTGGATT AGGAACC---
AA-IIb   331  ---------- -AGTGGCCAT CCTGTAAGC- -TGCTGGATT AGGAACC---
AA-IIc   334  ---------- -AGTGGCCAT CCTGTAAGC- -TGCTGGATT TGAAACC---
AA-IIa   331  ---------- -AGTGGCCAT CCTGTAAGC- -TGCTGGATT AGGAACC---
NA-IIa   331  ---------- -AGTGGCCAT CCTGTAAGC- -TGCTGGATT AGGAACC---
  CS-I   324  ---------- -AGTGGCCAT CCTGTGAGC- -CCCTGATTT AATGG-----
 Align   296  ---------- -mmmmmmmmm mmmmmmmmm- -mmmmmmmmm m---------

CA-IId   367  --CAG-TGGT ATTATCTCAG CCACAGA--- -TTTTCTGGT TGTAGTGGGC
NA-IIb   365  --CAG-TGGT ATTATCTCAG CCACAGA--- -TTTTCTGGT TGTAGTGGGC
AA-IIb   365  --CAG-TGGT ATTATCTCAG CCACAGA--- -TTTTCTGGT TGTAGTGGGC
AA-IIc   368  --CAG-TGGT ATTATCTCAG CCACAGA--- -TTTTCTGGT TGTAATGGGC
AA-IIa   365  --CAG-TGGT ATTATCTCAG CCACAGA--- -TTTTCTGGT TGTAATGGGC
NA-IIa   365  --CAG-TGGT ATTATCTCAG CCACAGA--- -TTTTCTGGT TGTAATGGGC
  CS-I   356  --CAGGCGGT ATCATCTCAG CCACATT--- -TCCTCTGGT AGTAGTGGGC
 Align   324  --mmm-mmmm mmmmmmmmmm mmmmm----- -mmmmmmmmm mmmmmmmmm

CA-IId   410  CAG------- ---------- ---------- ---------- ----------
NA-IIb   408  CAG------- ---------- ---------- ---------- ----------
AA-IIb   408  CAG------- ---------- ---------- ---------- ----------
AA-IIc   411  CAG------- ---------- ---------- ---------- ----------
AA-IIa   408  CAG------- ---------- ---------- ---------- ----------
NA-IIa   408  CAG------- ---------- ---------- ---------- ----------
  CS-I   400  CAG------- ---------- ---------- ---------- ----------
 Align   365  mmm------- ---------- ---------- ---------- ----------

CA-IId   413  ---------- ---------T TTTGA-AACT CGG----GGA ACATCT----
NA-IIb   411  ---------- ---------T TTTGA-AACT CGG----GGA ACATCT----
AA-IIb   411  ---------- ---------T TTTGA-AACT CGG----GGA ACATCT----
AA-IIc   414  ---------- ---------T TTTGA-AACT CGG----GGA ACATCT----
AA-IIa   411  ---------- ---------T TTTGA-AACT CGG----GGA ACATCT----
NA-IIa   411  ---------- ---------T TTTGA-AACT CGG----GGA ACATCT----
  CS-I   403  ---------- ---------A TTTAA-AACT CGA----GAA ACATCT----
 Align   368  ---------- ---------m mmmmm-mmmm mmm----mmm mmmmmm----

CA-IId   435  ---------- ---------- ---GTGACTT TCTTTCT--T -AA-CGCA--
NA-IIb   433  ---------- ---------- ---GTGACTT TCTTTCT--T -AA-CGCA--
AA-IIb   433  ---------- ---------- ---GTGACTT TCTTTCT--T -AA-CGCA--
AA-IIc   436  ---------- ---------- ---GTGACTT TCTTTCT--T -AA-CGCA--
AA-IIa   433  ---------- ---------- ---GTGACTT TCTTTCT--T -AA-CGCA--
NA-IIa   433  ---------- ---------- ---GTGACTT TCTTTCT--T -AA-CGCA--
  CS-I   425  ---------- ---------- ---GTGACTT TCTTTCT--T -TA-CGCA--
 Align   390  ---------- ---------- ---mmmmmmm mmmmmmm--m -mm-mmmm--
```

```
CA-IId   456  ---------- ---------- -GAGGAAGGT TATGGCAATA ACAGGTCTG-
NA-IIb   454  ---------- ---------- -GAGGAAGGT TATGGCAATA ACAGGTCTG-
AA-IIb   454  ---------- ---------- -GAGGAAGGT TATGGCAATA ACAGGTCTG-
AA-IIc   457  ---------- ---------- -GAGGAAGGT TATGGCAATA ACAGGTCTG-
AA-IIa   454  ---------- ---------- -GAGGAAGGT TATGGCAATA ACAGGTCTG-
NA-IIa   454  ---------- ---------- -GAGGAAGGT TATGGCAATA ACAGGTCTG-
  CS-I   446  ---------- ---------- -GAGGAAGGT TATGGCAATA ACAGGTCTG-
 Align   411  ---------- ---------- -mmmmmmmmmm mmmmmmmmmmm mmmmmmmmmm-

CA-IId   484  TGATGCCC-T TAG-ATGTCC AGGG-CTGCA CACGTACTAC ATTGA-TC-A
NA-IIb   482  TGATGCCC-T TAG-ATGTCC AGGG-CTGCA CACGTACTAC ATTGA-TC-A
AA-IIb   482  TGATGCCC-T TAG-ATGTCC AGGG-CTGCA CACGTACTAC ATTGA-TC-A
AA-IIc   485  TGATGCCC-T TAG-ATGTCT AGGG-CTGCA CACGTACTAC ATTGA-TC-A
AA-IIa   482  TGATGCCC-T TAG-ATGTCC AGGG-CTGCA CACGTACTAC ATTGA-TC-A
NA-IIa   482  TGATGCCC-T TAG-ATGTCC AGGG-CTGCA CACGTACTAC ATTGA-TC-A
  CS-I   474  TGATGCCC-T CAG-ATGTCC AGAG-CTGCA CACGTACTAC AGTGA-TC-G
 Align   439  mmmmmmmmm-m mmm-mmmmmm mmmm-mmmmmm mmmmmmmmmmm mmmmm-mm-m

CA-IId   529  ACT--CAGTA GGCGTCGT-G TTAGTTCTCC ---------- --------AA
NA-IIb   527  ACT--CAGTA GGCGTCGT-G TTAGTTCTCC ---------- --------AA
AA-IIb   527  ACT--CAGTA GGCGTCGT-G TTAGTTCTCC ---------- --------AA
AA-IIc   530  ACT--CAGTA GGCGTCGT-G CTTT--CTCC ---------- --------AA
AA-IIa   527  ACT--CAGTA GGCGTCGT-G TTTT--CTCC ---------- --------AA
NA-IIa   527  ACT--CAGTA GGCGTCGT-G TTTT--CTCC ---------- --------AA
  CS-I   519  ACT--CACTA AGTGTCGT-G TTTT--CTCC A--------- --------AA
 Align   484  mmm--mmmmm mmmmmmmm-m nnnn--mmmm ---------- --------mm

CA-IId   558  TAACGTATCT AGTGGACTTG GTGTCGGGTG TGTGGCCTCT GGTCATGCGC
NA-IIb   556  TAACGTATCT AGTGGACTTG GTGTCGGGTA TGTGGCCTC- GGTCATGTAC
AA-IIb   556  TAACGTATCT AGTGGACTTG GTGTCGGGTA TGTGGCCTC- GGTCATGTAC
AA-IIc   557  TAACGTATCT AGTGGACTTG GTGTCGGGTG CGTGGCCTC- GGTCGTGTAC
AA-IIa   554  TAACGTATCT AGTGGACTTG GTGTCGGGTG CGTGGCCTC- GGTCGTGTAC
NA-IIa   554  TAACGTATCT AGTGGACTTG GTGTCGGGTG CGTGGCCTC- GGTCGTGTAC
  CS-I   547  TAACGTATAC AGTGGACTTG GTGTCGGGTG C-TGGCCTCT GGTCATGTGC
 Align   511  mmmmmmmmmmm mmmmmmmmmmm mmmmmmmmmmm m-mmmmmmm- mmmmmmmmmmm

CA-IId   608  TTTGATTAC- ---------- ---------- ---------- ----------
NA-IIb   605  TTTGATTAC- ---------- ---------- ---------- ----------
AA-IIb   605  TTTGATTAC- ---------- ---------- ---------- ----------
AA-IIc   606  TTTGATTAC- ---------- ---------- ---------- ----------
AA-IIa   603  TTTGATTAC- ---------- ---------- ---------- ----------
NA-IIa   603  TTTGATTAC- ---------- ---------- ---------- ----------
  CS-I   596  TTTGATTAC- ---------- ---------- ---------- ----------
 Align   559  mmmmmmmmm- ---------- ---------- ---------- ----------

CA-IId   617  ---------- -----TGT-C ACTTT--AA- AC-----ACT G---------
NA-IIb   614  ---------- -----TGT-C ACTTT--AA- AC-----ACT G---------
AA-IIb   614  ---------- -----TGT-C ACTTT--AA- AC-----ACT G---------
AA-IIc   615  ---------- -----TGT-C ACTTT--AA- AC-----ACT G---------
AA-IIa   612  ---------- -----TGT-C ACTTT--AA- AC-----ACT G---------
NA-IIa   612  ---------- -----TGT-C ACTTT--AA- AC-----ACT G---------
  CS-I   605  ---------- -----TGT-C ACTTT--AA- AC-----ACT G---------
 Align   568  ---------- -----mmm-m mmmmm--mm- mm-----mmm m---------

CA-IId   634  ------GTCG T-TAGACTCG TGCAATCCAA TAG----AAG -AGGTGAATA
NA-IIb   631  ------GTCG T-TAGACTCG TGCAATCCAA TAG----AAG -AGGTGAATA
AA-IIb   631  ------GTCG T-TAGACTCG TGCAATCCAA TAG----AAG -AGGTGAATA
AA-IIc   632  ------GTCG T-TAGACTCG TGCAATCCAA TAG----AAG -AGGTGACTC
AA-IIa   629  ------GTCG T-TAGACTCG TGCAATCCAA TAG----AAG -AGGTGACTA
NA-IIa   629  ------GTCG T-TAGACTCG TGCAATCCAA TAG----AAG -AGGTGACTA
  CS-I   622  ------GTCG T-TAGACTCG TGCAA-GCAA TTC----A-G -AGC-AACGA
 Align   585  ------mmmm m-mmmmmmmm mmmmm-mmmm mmm----m-- -mmm------

CA-IId   672  AGTTTCCCAA AGT-----AC CGTGTTT--- ---------- ------AGC-
NA-IIb   669  AGTTTCCCAA AGT-----AC CGTGTTT--- ---------- ------AGC-
AA-IIb   669  AGTTTCCCAA AGT-----AC CGTGTTT--- ---------- ------AGC-
AA-IIc   670  AGTATCCCA- -GT-----AC CGTGTT---- ---------- ------AGTC
AA-IIa   667  AGTTTCCCA- -GT-----AC CGTGTT---- ---------- ------CGCC
NA-IIa   667  AGTTTCCCA- -GT-----AC CGTGTT---- ---------- ------CGCC
  CS-I   657  ATTG--CAAA C-------AC TCTTTG---- ---------- ------GGC-
 Align   614  ---------- --------mm mmmmm----- ---------- ----------

CA-IId   697  TCCCGTGTTC GT-AGAGA-- CCCC------ ---TCGCATT G---------
NA-IIb   694  TCCCGTGTTC GT-AGCGA-- CCCC------ ---TCGCAAA TG--------
AA-IIb   694  TCCCGTGTTC GT-AGCGA-- CCCC------ ---TCGCAAA TG--------
AA-IIc   693  TCCCGTGTTC GT-AGTGG-- CCCC------ ---GCGCATT G---------
AA-IIa   690  TCCCGTGT-C ---AGTAGGG CCCC------ ---TCGCATA TA--------
```

```
NA-IIa   690  TCCCGTGT-C ---AGTAGGG CCCC------ ---TCGCATA TA--------
  CS-I   677  ACTGTTT--- ---ACT---- CCC-------- ----G---TT T-----ACAG
 Align   621  ---------- ---------- ---------- ---------- ----------

CA-IId   726  ---------- ---------- -----CCAGT CGATGTGGGT CCT-------
NA-IIb   724  ---------- ---------- -----CCAGT CGATGTGGGT CCT-------
AA-IIb   724  ---------- ---------- -----CCAGT CGATGTGGGT CCT-------
AA-IIc   722  ---------- ---------- -----CCAGT CGTAGTGGGT CCT-------
AA-IIa   719  ---------- ---------- -----CCAGT CGAAGTGGGC CCT-------
NA-IIa   719  ---------- ---------- -----CCAGT CGAAGTGGGC CCT-------
  CS-I   698  AGGAGAGATA GAC-------- -TAA-CCACC TATTCTCGAC TCGC------
 Align   621  ---------- ---------- -----mmmmm mmmmmmmmmmm mmm-------

CA-IId   744  -------TTG TG--CTT--T TTTCAAAT-- ---CTTA--G TG-GGGACAG
NA-IIb   742  -------TTG TG--CTT--T TTTCAATT-- ---CTTA--G TG-GGGACAG
AA-IIb   742  -------TTG TG--CTT--T TTTCAATT-- ---CTTA--G TG-GGGACAG
AA-IIc   740  -------TTG TG--CTT--T TTTCAAAT-- ---CTTA--G TG-GGGACAG
AA-IIa   737  -------TTG TG--CTT--T TTTCAAAT-- ---CTTA--G TG-GGGACAG
NA-IIa   737  -------TTG TG--CTT--T TTTCAAAT-- ---CTTA--G TG-GGGACAG
  CS-I   733  -------CCG CC--CAT--C TTTCAATT-- ---CTTG--G TG-GGGACAG
 Align   639  ---------- ----mmm--m mmmmmmmmm-- ---mmmm--m mm-mmmmmmm

CA-IId   775  ACATCTGT-- TAACT-TTTT GTCTCGGTCC CAACCAGGAA TGCCTCGTAC
NA-IIb   773  ACATCTGT-- TAACT-TTTT GTCTCGGTCC CAACCAGGAA TGCCTCGTAC
AA-IIb   773  ACATCTGT-- TAACT-TTTT GTCTCGGTCC CAACCAGGAA TGCCTCGTAC
AA-IIc   771  ACATCTGT-- TAACT-TTTT GTCTCGGTCC CAACCAGGAA TGCCTCGTAC
AA-IIa   768  ACATCTGT-- TAACT-TTTT GTCTCGGTCC CAACCAGGAA TGCCTCGTAC
NA-IIa   768  ACATCTGT-- TAACT-TTTT GTCTCGGTCC CAACCAGGAA TGCCTCGTAC
  CS-I   764  TAGGTTGT-- TAACT-TTCT TACTCGGTCC TAACCAGGAA TGCCTCGTAC
 Align   665  mmmmmmmmm-- mmmmmm-mmmm mmmmmmmmmmm mmmmmmmmmmm mmmmmmmmmmm

CA-IId   822  AG-GTT---G GTTCACC-AT ACCACCTGGA ATTAGTCCCT GCCCTTTGTA
NA-IIb   820  AG-GTT---G GTTCACC-AT ACCACCTGGA ATTAGTCCCT GCCCTTTGTA
AA-IIb   820  AG-GTT---G GTTCACC-AT ACCACCTGGA ATTAGTCCCT GCCCTTTGTA
AA-IIc   818  AG-GTT---G GTTCACC-AT ACCACCTGGA ATTAGTCCCT GCCCTTTGTA
AA-IIa   815  AG-GTT---G GTTCACC-AT ACCACCTGGA ATTAGTCCCT GCCCTTTGTA
NA-IIa   815  AG-GTT---G GTTCACC-AT ACCACCTGGA ATTAGTCCCT GCCCTTTGTA
  CS-I   811  AG-GTT---G GTTCACC-AT ACCACCTGGA ATTAGTCCCT GCCCTTTGTA
 Align   712  mm-mmm---m mmmmmmm-mm mmmmmmmmmmm mmmmmmmmmmm mmmmmmmmmmm

CA-IId   867  CACACCGCCC GTCGCTTT-T ACCAA--TG- -AACTTCT-T TG-CGAGAGT
NA-IIb   865  CACACCGCCC GTCGCTTT-T ACCAA--TG- -AACTTCT-T TG-CGAGAGT
AA-IIb   865  CACACCGCCC GTCGCTTT-T ACCAA--TG- -AACTTCT-T TG-CGAGAGT
AA-IIc   863  CACACCGCCC GTCGCTTT-T ACCAA--TG- -AACTTCT-T TG-CGAGAGT
AA-IIa   860  CACACCGCCC GTCGCTTT-T ACCAA--TG- -AACTTCT-T TG-CGAGAGT
NA-IIa   860  CACACCGCCC GTCGCTTT-T ACCAA--TG- -AACTTCT-T TG-CGAGAGT
  CS-I   856  CACACCGCCC GTCGCTTT-T ACCAA--TG- -AACTACT-T TG-CGAGAGT
 Align   757  mmmmmmmmmm mmmmmmmm-m mmmmm--mm- -mmmmmm-m mm-mmmmmm

CA-IId   910  -GAGAGACTT A-----AAGA TA-------- ---------- ----------
NA-IIb   908  -AAGAGACTT G-----TAAT ---------- ---------- ----------
AA-IIb   908  -AAGAGACTT G-----TAAT ---------- ---------- ----------
AA-IIc   906  -GAGAGACTG A-----AATG T--------- ---------- ----------
AA-IIa   903  -GAGAGACTA A-----AGAT ---------- ---------- ----------
NA-IIa   903  -GAGAGACTA A-----AGAT ---------- ---------- ----------
  CS-I   899  -GTGGGACCT A-----TGAG CTTTCAAG-- ---------- ----------
 Align   800  -mmmmmmmm- ---------- ---------- ---------- ----------

CA-IId   926  -----ATCGC GG-AAC---T CAC-TCGAC- CGACG-GGAT TTAA-AGGAA
NA-IIb   922  ----AATCGT GG-AAC---T CAC-TCGAC- CGACG-GGAT TTAA-AGGAA
AA-IIb   922  ----AATCGT GG-AAC---T CAC-TCGAC- CGACG-GGAT TTAA-AGGAA
AA-IIc   921  ------GTGT GG-AAC---T CAC-TCGAC- CGACG-GGAC TTAA-AGGAA
AA-IIa   917  -----ATCGC GG-AAC---T CAC-TCGAC- CGACG-GGAC TTAA-AGGAA
NA-IIa   917  -----ATCGC GG-AAC---T CAC-TCGAC- CGACG-GGAC TTAA-AGGAA
  CS-I   921  -----TACAG GG-AAC---C CAT-TCGAC- CAACG-GAGT TTAA-AGGAA
 Align   808  ---------- mm-mmm---m mmm-mmmmm- mmmmm-mmmm mmmm-mmmmm

CA-IId   963  AA-AGAA-G- TCGT-AAC
NA-IIb   960  AA-AGAA-G- TCGT-AAC
AA-IIb   960  AA-AGAA-G- TCGT-AAC
AA-IIc   957  AA-AGAA-G- TCGT-AAC
AA-IIa   954  AA-AGAA-G- TCGT-AAC
NA-IIa   954  AA-AGAA-G- TCGT-AAC
  CS-I   958  AA-AGAA-G- TCGT-AAC
 Align   840  mm-mmmm-m- mmmm-mmm
```

A2.1.2. *Globigerina bulloides* Type II 935 bp molecular phylogeny

```
CA-IId      1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
NA-IIb      1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
AA-IIb      1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
AA-IIc      1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
AA-IIa      1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
NA-IIa      1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TTGACTCAAC
 Align      1  mmm--mmmmm mm--mmmmm- --mmmmmm-m mmmmmmmmmm mmmmmmmmmm

CA-IId     43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACAC-AGTGA G--GATTG-A
NA-IIb     43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACGC-AGTGA G--GATTG-A
AA-IIb     43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACGC-AGTGA G--GATTG-A
AA-IIc     43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACAC-AGTGA G--GATTG-A
AA-IIa     43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACGC-AGTGA G--GATTG-A
NA-IIa     43  GCGGA-AAAG CTTATCTGG- -----TCCGG ACGC-AGTGA G--GATTG-A
 Align     43  mmmmm-mmmm mmmmmmmmm- -----mmmmm mmmm-mmmmm m--mmmmm-m

CA-IId     82  CAGACAGTT- ------AGA- CAGAAGT-GA -GGTTTT--G GTAA---CAA
NA-IIb     82  CAGACAGTT- ------AGA- CAGAAGT--- -GGTTTA--G GTAA---CAA
AA-IIb     82  CAGACAGTT- ------AGA- CAGAAGT--- -GGTTTA--G GTAA---CAA
AA-IIc     82  CAGACAGTT- -------GG- CAGGAGT--- -GGTTCTT-G GTAAAAACAA
AA-IIa     82  CAGACAGTT- --------T- CAGGAGT--- -GGTTCTT-G GTAAA--CAA
NA-IIa     82  CAGACAGTT- --------T- CAGGAGT--- -GGTTCTT-G GTAAA--CAA
 Align     82  mmmmmmmmm- ---------- mmmmmmmm--- -mmm-----m mmmmm--mmm

CA-IId    117  ---TTGAGAG AGTTGAA-GT ---------- ---------T CTTTCATGA-
NA-IIb    115  ---TTGAGAG AGTTGAA-GT ---------- ---------T CTTTCATGA-
AA-IIb    115  ---TTGAGAG AGTTGAA-GT ---------- ---------T CTTTCATGA-
AA-IIc    118  ---ATGAGAG AGTTGAA-GT ---------- ---------T CTTTCATGA-
AA-IIa    115  ---TTGAGAG AGTTGAA-GT ---------- ---------T CTTTCATGA-
NA-IIa    115  ---TTGAGAG AGTTGAA-GT ---------- ---------T CTTTCATGA-
 Align    110  ---mmmmmmm mmmmmmm-mm ---------- ---------m mmmmmmmmm-

CA-IId    143  -----TCTTG TGAGAG---- -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT
NA-IIb    141  -----TCTTG TGAGAG---- -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT
AA-IIb    141  -----TCTTG TGAGAG---- -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT
AA-IIc    144  -----TCTTG TGAGAG---- -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT
AA-IIa    141  -----TCTTG TGAGAG---- -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT
NA-IIa    141  -----TCTTG TGAGAG---- -GTGGTG-CA TGG-CCGTT- CTTAGTTCGT
 Align    136  -----mmmmm mmmmmm---- -mmmmmm-mm mmm-mmmmm- mmmmmmmmmm

CA-IId    180  GTA-GTGATA -TGTC--TGC CT-AATCGCG TCAC------ -GATAACCTA
NA-IIb    178  GTA-GTGATA -TGTC--TGC CT-AATCGCG TCAC------ -GATAACCTA
AA-IIb    178  GTA-GTGATA -TGTC--TGC CT-AATCGCG TCAC------ -GATAACCTA
AA-IIc    181  GTA-GTGATA -TGTC--TGC CT-AATCGCG TCAC------ -GATAACCTA
AA-IIa    178  GTA-GTGATA -TGTC--TGC CT-AATCGCG TCAC------ -GATAACCTA
NA-IIa    178  GTA-GTGATA -TGTC--TGC CT-AATCGCG TCAC------ -GATAACCTA
 Align    173  mmm-mmmmmm -mmmm--mmm mm-mmmmmmm mmmm------ -mmmmmmmmm

CA-IId    218  TTGGTCGACA ACCCAATTAT C-ACAACTGC AGCATAACTC CCC-TTGGGT
NA-IIb    216  TTGGTCGACA ACCCAATTAT C-ACAACTGC AGCATAACTC CCC-TTGGGT
AA-IIb    216  TTGGTCGACA ACCCAATTAT C-ACAACTGC AGCATAACTC CCC-TTGGGT
AA-IIc    219  TTGGTCGACA ACCCAATTAT C-ACAACTGC AGCATAACTC CCC-TTGGGT
AA-IIa    216  TTGGTCGACA ACCCAATTAT C-ACAACTGC AGCATAACTC CCC-TTGGGT
NA-IIa    216  TTGGTCGACA ACCCAATTAT C-ACAACTGC AGCATAACTC CCC-TTGGGT
 Align    211  mmmmmmmmmm mmmmmmmmmm m-mmmmmmmm mmmmmmnnnn nnn-mmmmm

CA-IId    266  GGGCGAGGCT CTGTGTAGGA TAGAC----- ---------- ----------
NA-IIb    264  GGGCGAGGCT CTGTGTAGGA TAGAC----- ---------- ----------
AA-IIb    264  GGGCGAGGCT CTGTGTAGGA TAGAC----- ---------- ----------
AA-IIc    267  GGGCGAGGCT CTGTGTAGGA TAGAC----- ---------- ----------
AA-IIa    264  GGGCGAGGCT CTGTGTAGGA TAGAC----- ---------- ----------
NA-IIa    264  GGGCGAGGCT CTGTGTAGGA TAGAC----- ---------- ----------
 Align    259  mmmmmmmmmm mmmmmmmmmm mmmmm----- ---------- ----------

CA-IId    291  ---------- ---------- ---------- ---------- ---------C
NA-IIb    289  ---------- ---------- ---------- ---------- ---------C
AA-IIb    289  ---------- ---------- ---------- ---------- ---------C
AA-IIc    292  ---------- ---------- ---------- ---------- ---------C
AA-IIa    289  ---------- ---------- ---------- ---------- ---------C
NA-IIa    289  ---------- ---------- ---------- ---------- ---------C
 Align    284  ---------- ---------- ---------- ---------- ---------m

CA-IId    292  TCTGAACAGT ACGCAACGAA CGCGATCGT- AA----TCCC TTGT--TG--
```

```
NA-IIb   290  TCTGAACAGT ACGCAACGAA CGCGATCGT- AA----TCCC TTGT--TG--
AA-IIb   290  TCTGAACAGT ACGCAACGAA CGCGATCGT- AA----TCCC TTGT--TG--
AA-IIc   293  TCTGAACAGT ACGCAACGAA CGCGATCGT- AA----TCCC TTGT--TG--
AA-IIa   290  TCTGAACAGT ACGCAACGAA CGCGATCGT- AA----TCCC TTGT--TG--
NA-IIa   290  TCTGAACAGT ACGCAACGAA CGCGATCGT- AA----TCCC TTGT--TG--
 Align   285  mmmmmmmmmmm mmmmmmmmmmmm mmmmmmmmmm- mm----mmmm mmmm--mm--

CA-IId   333  ---------- -----AGTGG CCATCCTGTA AGC--TGCTG GATTAGGAAC
NA-IIb   331  ---------- -----AGTGG CCATCCTGTA AGC--TGCTG GATTAGGAAC
AA-IIb   331  ---------- -----AGTGG CCATCCTGTA AGC--TGCTG GATTAGGAAC
AA-IIc   334  ---------- -----AGTGG CCATCCTGTA AGC--TGCTG GATTTGAAAC
AA-IIa   331  ---------- -----AGTGG CCATCCTGTA AGC--TGCTG GATTAGGAAC
NA-IIa   331  ---------- -----AGTGG CCATCCTGTA AGC--TGCTG GATTAGGAAC
 Align   326  ---------- -----mmmmm mmmmmmmmmmm mmm--mmmmm mmmmmmmmmm

CA-IId   366  C-----CAG- TGGTATTATC TCAGCCACAG A----TTTTC TGGTTGTAGT
NA-IIb   364  C-----CAG- TGGTATTATC TCAGCCACAG A----TTTTC TGGTTGTAGT
AA-IIb   364  C-----CAG- TGGTATTATC TCAGCCACAG A----TTTTC TGGTTGTAGT
AA-IIc   367  C-----CAG- TGGTATTATC TCAGCCACAG A----TTTTC TGGTTGTAAT
AA-IIa   364  C-----CAG- TGGTATTATC TCAGCCACAG A----TTTTC TGGTTGTAAT
NA-IIa   364  C-----CAG- TGGTATTATC TCAGCCACAG A----TTTTC TGGTTGTAAT
 Align   359  m-----mmm- mmmmmmmmmm mmmmmmmmmm m----mmmmm mmmmmmmmmm

CA-IId   406  GGGCCAG--- ---------- ---------- ---------- ----------
NA-IIb   404  GGGCCAG--- ---------- ---------- ---------- ----------
AA-IIb   404  GGGCCAG--- ---------- ---------- ---------- ----------
AA-IIc   407  GGGCCAG--- ---------- ---------- ---------- ----------
AA-IIa   404  GGGCCAG--- ---------- ---------- ---------- ----------
NA-IIa   404  GGGCCAG--- ---------- ---------- ---------- ----------
 Align   399  mmmmmmm--- ---------- ---------- ---------- ----------

CA-IId   413  ---------- ---------- ---TTTTGA- AACTCGG--- -GGAACATCT
NA-IIb   411  ---------- ---------- ---TTTTGA- AACTCGG--- -GGAACATCT
AA-IIb   411  ---------- ---------- ---TTTTGA- AACTCGG--- -GGAACATCT
AA-IIc   414  ---------- ---------- ---TTTTGA- AACTCGG--- -GGAACATCT
AA-IIa   411  ---------- ---------- ---TTTTGA- AACTCGG--- -GGAACATCT
NA-IIa   411  ---------- ---------- ---TTTTGA- AACTCGG--- -GGAACATCT
 Align   406  ---------- ---------- ---mmmmmm- mmmmmmmm--- -mmmmmmmmm

CA-IId   435  ---------- ---------- -------GTG ACTTTCTTTC T--T-AA-CG
NA-IIb   433  ---------- ---------- -------GTG ACTTTCTTTC T--T-AA-CG
AA-IIb   433  ---------- ---------- -------GTG ACTTTCTTTC T--T-AA-CG
AA-IIc   436  ---------- ---------- -------GTG ACTTTCTTTC T--T-AA-CG
AA-IIa   433  ---------- ---------- -------GTG ACTTTCTTTC T--T-AA-CG
NA-IIa   433  ---------- ---------- -------GTG ACTTTCTTTC T--T-AA-CG
 Align   428  ---------- ---------- -------mmm mmmmmmmmmm m--m-mm-mm

CA-IId   454  CA-------- ---------- -----GAGGA AGGTTATGGC AATAACAGGT
NA-IIb   452  CA-------- ---------- -----GAGGA AGGTTATGGC AATAACAGGT
AA-IIb   452  CA-------- ---------- -----GAGGA AGGTTATGGC AATAACAGGT
AA-IIc   455  CA-------- ---------- -----GAGGA AGGTTATGGC AATAACAGGT
AA-IIa   452  CA-------- ---------- -----GAGGA AGGTTATGGC AATAACAGGT
NA-IIa   452  CA-------- ---------- -----GAGGA AGGTTATGGC AATAACAGGT
 Align   447  mm-------- ---------- -----mmmmm mmmmmmmmmm mmmmmmmmmm

CA-IId   481  CTG-TGATGC CC-TTAG-AT GTCCAGGG-C TGCACACGTA CTACATTGA-
NA-IIb   479  CTG-TGATGC CC-TTAG-AT GTCCAGGG-C TGCACACGTA CTACATTGA-
AA-IIb   479  CTG-TGATGC CC-TTAG-AT GTCCAGGG-C TGCACACGTA CTACATTGA-
AA-IIc   482  CTG-TGATGC CC-TTAG-AT GTCTAGGG-C TGCACACGTA CTACATTGA-
AA-IIa   479  CTG-TGATGC CC-TTAG-AT GTCCAGGG-C TGCACACGTA CTACATTGA-
NA-IIa   479  CTG-TGATGC CC-TTAG-AT GTCCAGGG-C TGCACACGTA CTACATTGA-
 Align   474  mmm-mmmmmm mm-mmmm-mm mmmmmmmmm-m mmmmmmmmmmm mmmmmmmmmm-

CA-IId   526  TC-AACT--C AGTAGGCGTC GT-GTTAGTT CTCC------ ----------
NA-IIb   524  TC-AACT--C AGTAGGCGTC GT-GTTAGTT CTCC------ ----------
AA-IIb   524  TC-AACT--C AGTAGGCGTC GT-GTTAGTT CTCC------ ----------
AA-IIc   527  TC-AACT--C AGTAGGCGTC GT-GCT--TT CTCC------ ----------
AA-IIa   524  TC-AACT--C AGTAGGCGTC GT-GTT--TT CTCC------ ----------
NA-IIa   524  TC-AACT--C AGTAGGCGTC GT-GTT--TT CTCC------ ----------
 Align   519  mm-mmmm--m mmmmmmmmmm mm-mmm--mm mmmm------ ----------

CA-IId   556  --AATAACGT ATCTAGTGGA CTTGGTGTCG GGTGTGTGGC CTCTGGTCAT
NA-IIb   554  --AATAACGT ATCTAGTGGA CTTGGTGTCG GGTATGTGGC CTC-GGTCAT
AA-IIb   554  --AATAACGT ATCTAGTGGA CTTGGTGTCG GGTATGTGGC CTC-GGTCAT
AA-IIc   555  --AATAACGT ATCTAGTGGA CTTGGTGTCG GGTGCGTGGC CTC-GGTCGT
AA-IIa   552  --AATAACGT ATCTAGTGGA CTTGGTGTCG GGTGCGTGGC CTC-GGTCGT
NA-IIa   552  --AATAACGT ATCTAGTGGA CTTGGTGTCG GGTGCGTGGC CTC-GGTCGT
```

```
Align   547  --mmmmmmmmm mmmmmmmmmmm mmmmmmmmmmm mmmmmmmmmmm mmm-mmmmmm

CA-IId  604  GCGCTTTGAT TAC------- ---------- ---------- ----------
NA-IIb  601  GTACTTTGAT TAC------- ---------- ---------- ----------
AA-IIb  601  GTACTTTGAT TAC------- ---------- ---------- ----------
AA-IIc  602  GTACTTTGAT TAC------- ---------- ---------- ----------
AA-IIa  599  GTACTTTGAT TAC------- ---------- ---------- ----------
NA-IIa  599  GTACTTTGAT TAC------- ---------- ---------- ----------
Align   594  mmmmmmmmmm mmm------- ---------- ---------- ----------

CA-IId  617  ---------- ---------T GTCACTTT-- AA-AC----- ACTG------
NA-IIb  614  ---------- ---------T GTCACTTT-- AA-AC----- ACTG------
AA-IIb  614  ---------- ---------T GTCACTTT-- AA-AC----- ACTG------
AA-IIc  615  ---------- ---------T GTCACTTT-- AA-AC----- ACTG------
AA-IIa  612  ---------- ---------T GTCACTTT-- AA-AC----- ACTG------
NA-IIa  612  ---------- ---------T GTCACTTT-- AA-AC----- ACTG------
Align   607  ---------- ---------m mmmmmmmm-- mm-mm----- mmmm------

CA-IId  634  ---------G TCGT-TAGAC TCGTGCAATC CAATAG---- AAG-AGGTGA
NA-IIb  631  ---------G TCGT-TAGAC TCGTGCAATC CAATAG---- AAG-AGGTGA
AA-IIb  631  ---------G TCGT-TAGAC TCGTGCAATC CAATAG---- AAG-AGGTGA
AA-IIc  632  ---------G TCGT-TAGAC TCGTGCAATC CAATAG---- AAG-AGGTGA
AA-IIa  629  ---------G TCGT-TAGAC TCGTGCAATC CAATAG---- AAG-AGGTGA
NA-IIa  629  ---------G TCGT-TAGAC TCGTGCAATC CAATAG---- AAG-AGGTGA
Align   624  ---------m mmmm-mmmmm mmmmmmmmmm mmmmmm---- mmm-mmmmmm

CA-IId  669  ATAAGTTTCC CAAAGT---- -ACCGTGTTT ---------- ---------A
NA-IIb  666  ATAAGTTTCC CAAAGT---- -ACCGTGTTT ---------- ---------A
AA-IIb  666  ATAAGTTTCC CAAAGT---- -ACCGTGTTT ---------- ---------A
AA-IIc  667  CTCAGTATCC CA--GT---- -ACCGTGTT- ---------- ---------A
AA-IIa  664  CTAAGTTTCC CA--GT---- -ACCGTGTT- ---------- ---------C
NA-IIa  664  CTAAGTTTCC CA--GT---- -ACCGTGTT- ---------- ---------C
Align   659  mmmmmmmmmm mm--mm---- -mmmmmmmm- ---------- ----------

CA-IId  695  GC--TCCCGT GT-TCGT-AG AGA---CCCC ---------T CGCA--TTG-
NA-IIb  692  GC--TCCCGT GT-TCGT-AG CGA---CCCC ---------T CGCA--AATG
AA-IIb  692  GC--TCCCGT GT-TCGT-AG CGA---CCCC ---------T CGCA--AATG
AA-IIc  690  GTC-TCCCGT GT-TCGT-AG TGG---CCCC ---------G CGCA--TTG-
AA-IIa  687  GCC-TCCCGT GT--C---AG TAGGG-CCCC ---------T CGCA--TATA
NA-IIa  687  GCC-TCCCGT GT--C---AG TAGGG-CCCC ---------T CGCA--TATA
Align   681  ----mmmmmm mm-------- ------mmmm ---------m mmmm------

CA-IId  726  ---------- ---------- ---------- ---CCAGTCG ATGTGGGTCC
NA-IIb  724  ---------- ---------- ---------- ---CCAGTCG ATGTGGGTCC
AA-IIb  724  ---------- ---------- ---------- ---CCAGTCG ATGTGGGTCC
AA-IIc  722  ---------- ---------- ---------- ---CCAGTCG TAGTGGGTCC
AA-IIa  719  ---------- ---------- ---------- ---CCAGTCG AAGTGGGCCC
NA-IIa  719  ---------- ---------- ---------- ---CCAGTCG AAGTGGGCCC
Align   698  ---------- ---------- ---------- ---mmmmmmm mmmmmmmmmm

CA-IId  743  T--------- ---------- ---------- ---------- ----------
NA-IIb  741  T--------- ---------- ---------- ---------- ----------
AA-IIb  741  T--------- ---------- ---------- ---------- ----------
AA-IIc  739  T--------- ---------- ---------- ---------- ----------
AA-IIa  736  T--------- ---------- ---------- ---------- ----------
NA-IIa  736  T--------- ---------- ---------- ---------- ----------
Align   715  m--------- ---------- ---------- ---------- ----------

CA-IId  744  ---------- -----TTGTG --CTT--TTT TCAAAT---- -CTTA--GTG
NA-IIb  742  ---------- -----TTGTG --CTT--TTT TCAATT---- -CTTA--GTG
AA-IIb  742  ---------- -----TTGTG --CTT--TTT TCAATT---- -CTTA--GTG
AA-IIc  740  ---------- -----TTGTG --CTT--TTT TCAAAT---- -CTTA--GTG
AA-IIa  737  ---------- -----TTGTG --CTT--TTT TCAAAT---- -CTTA--GTG
NA-IIa  737  ---------- -----TTGTG --CTT--TTT TCAAAT---- -CTTA--GTG
Align   716  ---------- -----mmmmm --mmm--mmm mmmmmmm---- -mmmm--mmm

CA-IId  768  -GGGACAGAC ATCTGT--TA ACT-TTTTGT CTCGGTCCCA ACCAGGAATG
NA-IIb  766  -GGGACAGAC ATCTGT--TA ACT-TTTTGT CTCGGTCCCA ACCAGGAATG
AA-IIb  766  -GGGACAGAC ATCTGT--TA ACT-TTTTGT CTCGGTCCCA ACCAGGAATG
AA-IIc  764  -GGGACAGAC ATCTGT--TA ACT-TTTTGT CTCGGTCCCA ACCAGGAATG
AA-IIa  761  -GGGACAGAC ATCTGT--TA ACT-TTTTGT CTCGGTCCCA ACCAGGAATG
NA-IIa  761  -GGGACAGAC ATCTGT--TA ACT-TTTTGT CTCGGTCCCA ACCAGGAATG
Align   740  -mmmmmmmmm mmmmmm--mm mmm-mmmmmm mmmmmmmmmm mmmmmmmmmm

CA-IId  814  CCTCGTACAG -GTT---GGT TCACC-ATAC CACCTGGAAT TAGTCCCTGC
NA-IIb  812  CCTCGTACAG -GTT---GGT TCACC-ATAC CACCTGGAAT TAGTCCCTGC
AA-IIb  812  CCTCGTACAG -GTT---GGT TCACC-ATAC CACCTGGAAT TAGTCCCTGC
```

```
AA-IIc   810  CCTCGTACAG -GTT---GGT TCACC-ATAC CACCTGGAAT TAGTCCCTGC
AA-IIa   807  CCTCGTACAG -GTT---GGT TCACC-ATAC CACCTGGAAT TAGTCCCTGC
NA-IIa   807  CCTCGTACAG -GTT---GGT TCACC-ATAC CACCTGGAAT TAGTCCCTGC
 Align   786  mmmmmmmmmmm -mmm---mmm mmmmm-mmmm mmmmmmmmmmm mmmmmmmmmmm

CA-IId   859  CCTTTGTACA CACCGCCCGT CGCTTT-TAC CAA--TG--A ACTTCT-TTG
NA-IIb   857  CCTTTGTACA CACCGCCCGT CGCTTT-TAC CAA--TG--A ACTTCT-TTG
AA-IIb   857  CCTTTGTACA CACCGCCCGT CGCTTT-TAC CAA--TG--A ACTTCT-TTG
AA-IIc   855  CCTTTGTACA CACCGCCCGT CGCTTT-TAC CAA--TG--A ACTTCT-TTG
AA-IIa   852  CCTTTGTACA CACCGCCCGT CGCTTT-TAC CAA--TG--A ACTTCT-TTG
NA-IIa   852  CCTTTGTACA CACCGCCCGT CGCTTT-TAC CAA--TG--A ACTTCT-TTG
 Align   831  mmmmmmmmmmm mmmmmmmmmmm mmmmmm-mmm mmm--mm--m mmmmmm-mmm

CA-IId   903  -CGAGAGT-G AGAGACTTA- ----AAGATA AT-------- ----------
NA-IIb   901  -CGAGAGT-A AGAGACTTG- -----TAATA AT-------- ----------
AA-IIb   901  -CGAGAGT-A AGAGACTTG- -----TAATA AT-------- ----------
AA-IIc   899  -CGAGAGT-G AGAGACTGA- -----AATGT GT-------- ----------
AA-IIa   896  -CGAGAGT-G AGAGACTAA- ------AGAT AT-------- ----------
NA-IIa   896  -CGAGAGT-G AGAGACTAA- ------AGAT AT-------- ----------
 Align   875  -mmmmmmmm-m mmmmmmmm--- ---------- ---------- ----------

CA-IId   928  ---------- ------CGCG G-AAC---TC AC-TCGAC-C GACG-GGATT
NA-IIb   925  ---------- ------CGTG G-AAC---TC AC-TCGAC-C GACG-GGATT
AA-IIb   925  ---------- ------CGTG G-AAC---TC AC-TCGAC-C GACG-GGATT
AA-IIc   923  ---------- -------GTG G-AAC---TC AC-TCGAC-C GACG-GGACT
AA-IIa   919  ---------- ------CGCG G-AAC---TC AC-TCGAC-C GACG-GGACT
NA-IIa   919  ---------- ------CGCG G-AAC---TC AC-TCGAC-C GACG-GGACT
 Align   890  ---------- ---------m m-mmm---mm mm-mmmmm-m mmmm-mmmmm

CA-IId   955  TAA-AGGAAA A-AGAA-G-T CGT-AAC
NA-IIb   952  TAA-AGGAAA A-AGAA-G-T CGT-AAC
AA-IIb   952  TAA-AGGAAA A-AGAA-G-T CGT-AAC
AA-IIc   949  TAA-AGGAAA A-AGAA-G-T CGT-AAC
AA-IIa   946  TAA-AGGAAA A-AGAA-G-T CGT-AAC
NA-IIa   946  TAA-AGGAAA A-AGAA-G-T CGT-AAC
 Align   914  mmm-mmmmmm m-mmmm-m-m mmm-mmm
```

A2.1.3. *Turborotailta quinqueloba* 762 bp molecular phylogeny

```
NA-IIa     1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TCGACTCAAC
AA-IIa     1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TCGACTCAAC
NA-IIb     1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TCGACTCAAC
AA-IIc     1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TCGACTCAAC
  CS-I     1  GCA--CCACA AG--AGCGT- --GGAGTA-T GTGGCTTAAT TCGACTCAAC
 Align     1  mmm--mmmmm mm--mmmmm- --mmmmmm-m mmmmmmmmmmm mmmmmmmmmmm

NA-IIa    43  GCGCA-ACAA TTTACTTGG- -----TCCGA ACGC-TTTGG G--G-TTG-A
AA-IIa    43  GCGCA-ACAA TTTACTTGG- -----TCCGA ACGC-TTTGG G--GATTG-A
NA-IIb    43  GCGCA-ACAA TTTACTTGG- -----TCCGA ACGC-TTTGA G--GATTG-A
AA-IIc    43  GCGCA-ACAA TTTACTTGG- -----TCCGA ACGC-TTTGA G--GATTG-A
  CS-I    43  GCGCA-ATAA CTTACTTGG- -----TCCGA ACGC-TTTGA G--GATTG-A
 Align    43  mmmmm-mmmm mmmmmmmmm- -----mmmmm mmmm-mmmmm m--mmmmm-m

NA-IIa    81  CAGTTATTG- ------TATA GTTCTGATAT GAGAGGTCTT TGTAGTCAAC
AA-IIa    82  CAGTTATTG- ------TATA GTTCTGATAT GAGAGGTCTT TGTAGTCAAC
NA-IIb    82  CAGTTATTG- ------TATA GTTCTGATAT GGGTGGTATT TGTAGTCAAC
AA-IIc    82  CAGTTATTG- ------TATA GTTCTGATAT GGGTGGTATT TGTAGTCAAC
  CS-I    82  CAGTTTCTG- ------GTAC AATGTGCGGC TTGTTTGTTA CAACTACGAA
 Align    82  mmmmmmmmm- ---------- ---------- ---------- ----------

NA-IIa   124  GTGTAGGTAG TTGTA----- ---------- ---------- ----------
AA-IIa   125  GTGTAGGTAG TTGTA----- ---------- ---------- ----------
NA-IIb   125  GTGTAGGTAG TTGTA----- ---------- ---------- ----------
AA-IIc   125  GTGTAGGTAG TTGTA----- ---------- ---------- ----------
  CS-I   125  TGATTCGAAT TGTTGTAAAT ATGACTTGGC CGGCCTTCGG GTGTCTTGGA
 Align    91  ---------- ---------- ---------- ---------- ----------

NA-IIa   139  ---------- ---------- ---------- --TAT-TAAA TATGAAA-GT
AA-IIa   140  ---------- ---------- ---------- --TAT-TAAA TATGAAA-GT
```

```
NA-IIb   140  ---------- ---------- ---------- --TAT-TAAA TATGAAA-GT
AA-IIc   140  ---------- ---------- ---------- --TAT-TAAA TATGAAA-GT
  CS-I   175  TCGGCGTATC AGGTCGTACA TTG------- --TAT-TCAA TGAGAAA-GT
 Align    91  ---------- ---------- ---------- --mmm-mmmm mmmmmmmm-mm

NA-IIa   155  ---------- ---------T CTTTTATGA- -----TTATG TGATAG----
AA-IIa   156  ---------- ---------T CTTTTATGA- -----TTATG TGATAG----
NA-IIb   156  ---------- ---------T CTTTTATGA- -----TTATG TGATAG----
AA-IIc   156  ---------- ---------T CTTTTATGA- -----TTATG TGATAG----
  CS-I   214  ---------- ---------T CTTTTATGA- -----TTATG TGGTAG----
 Align   107  ---------- ---------m mmmmmmmmmm- -----mmmmm mmmmmm----

NA-IIa   176  -GTGGTG-CA TGG-CCGTC- CTTAATTCGT GGA-GTGATT -TGTC--TGC
AA-IIa   177  -GTGGTG-CA TGG-CCGTC- CTTAATTCGT GGA-GTGATT -TGTC--TGC
NA-IIb   177  -GTGGTG-CA TGG-CCGTC- CTTAATTCGT GGA-GTGATT -TGTC--TGC
AA-IIc   177  -GTGGTG-CA TGG-CCGTC- CTTAATTCGT GGA-GTGATT -TGTC--TGC
  CS-I   235  -GTGGTG-CA TGG-CCGTC- TTTAATTCGT GGA-GTGATC -TGTC--TGC
 Align   128  -mmmmmm-mm mmm-mmmmm- mmmmmmmmmm mmm-mmmmmm -mmmm--mmm

NA-IIa   218  TT-AATTGCG CATT------ -GCAAATTGT AATTGATCTT TTACCAGTCT
AA-IIa   219  TT-AATTGCG CATT------ -GCAAATTGT AATTGATCTT TTACCAGTCT
NA-IIb   219  TT-AATTGCG CATT------ -GCAAATTGT AATTGATCTT TTACCAGTCT
AA-IIc   219  TT-AATTGCG CATT------ -GCAAATTGT AATTGATCTT TTACCAGTCT
  CS-I   277  TT-AATTGCG CATT------ -GCAAATTGT AATTGATCTT TGAACAGATC
 Align   170  mm-mmmmmmm mmmm------ -mmmmmmmmm mmmmmmmmmm mmmmmm---

NA-IIa   260  CGTTCCGATT GATGTTGTGA ATATTGTAGA ATATTGTTGT CGTCGAGTCA
AA-IIa   261  CGTTCCGATT GATGTTGTGA ATATTGTAGA ATATTGTTGT CGTCGAGTCA
NA-IIb   261  CGTTCTGATT TATGTTGT-A ATATTGTAGA ATATTGTTGT CGTCGAGTCA
AA-IIc   261  CGTTCTGATT TATGTTGT-A ATATTGTAGA ATATTGTTGT CGTCGAGTCA
  CS-I   319  CGTCTTTTT- -ATGTTGAAT AGTACGTACA GTCAGACCCG GTTGGGCATG
 Align   209  ---------- ---------- ---------- ---------- ----------

NA-IIa   310  TAATTGCCAC TACGTCTGGT CCTTCGGGCA GAATAGTCCT TATGCTTGTC
AA-IIa   311  TAATTGCCAC TACGTCTGGT CCTTCGGGCA GAATAGTCCT TATGCTTGTC
NA-IIb   310  TTATTGCCAC TACGTCTGGT CCTTTGGGGG CAGAATAGTG TTGTATATTT
AA-IIc   310  TTATTGCCAC TACGTCTGGT CCTTTGGG-- CAGAATAGTG TTGTATATTT
  CS-I   367  TACGGAAAGG TCACATATAA AGTACGGTCT GCTAAAGATA TGAATA----
 Align   209  ---------- ---------- ---------- ---------- ----------

NA-IIa   360  T--------AC ACAAAATTCT CCACAATGTT GTAGCCATAC TTGATTGTAT
AA-IIa   361  T--------AC ACAAAATTCT CCACAATGTT GTAGCCATAC TTGATTGTAT
NA-IIb   360  TGTCTTGTCA ACACAAAATT CTTCACAATA TTGTAGCCGT ACTTGATTGT
AA-IIc   358  TGTCTTGTCA ACACAAAATT CTTCACAATA TTGTAGCCGT ACTTGATTGT
  CS-I   413  ---------- ---------- ---------- ---------- ----------
 Align   209  ---------- ---------- ---------- ---------- ----------

NA-IIa   403  GCGCTATTAA TGTCTATGTG TGACACATTC GTGGTTCAAG ACCAGTTCGG
AA-IIa   404  GCGCTATTAA TGTCTATGTG TGACACATTC GTGGTTCAAG ACCAGTTCGG
NA-IIb   410  ATGCGCTATT CATGTATATG TG--GTTAGT ------CAAC ACCAGTTCGG
AA-IIc   408  ATGCGCTATT CATGTATATG TG--GTTAGT ------CAAC ACCAGTTCGG
  CS-I   413  ---------- ---------- ---------- ---------- ----------
 Align   209  ---------- ---------- ---------- ---------- ----------

NA-IIa   453  CTGCTTGCAC ACATTGCCAT CTCAACATGA TTGGTCAGGC TGCTAGAGAC
AA-IIa   454  CTGCTTGCAC ACATTGCCAT CTCAACATGA TTGGTCAGGC TGCTAGAGAC
NA-IIb   452  CTGCTTGCAC ACATCGCCAT CTCAACATGA TCAGTCAGGC TGCTAGAGAC
AA-IIc   450  CTGCTTGCAC ACATCGCCAT CTCAACATGA TCAGTCAGGC TGCTAGAGAC
  CS-I   413  ---------- ---------- ---------- ---------- ----------
 Align   209  ---------- ---------- ---------- ---------- ----------

NA-IIa   503  CTTGTG---- TCTTGCCCCA AAGGTTG--- ------ATTA ACACCGTGAG
AA-IIa   504  CTTGTG---- TCTTGCCCCA AAGGTTG--- ------ATTA ACACCGTGAG
NA-IIb   502  CTTGTG---- TCTTGCCCCA AAGGTTG--- ------ATTA ACACCGTGAG
AA-IIc   500  CTTGTG---- TCTTGCCCCA AAGGTTG--- ------ATTA ACACCGTGAG
  CS-I   413  ---------- TCTTGCCCTA AAGATTG--- ------ATTA ACACCGTGAG
 Align   209  ---------- mmmmmmmmmm mmmmmmmm--- ------mmmm mmmmmmmmmm

NA-IIa   540  TGCAACGAGT GAGATTGC-A A-----GTCT TTG--TT--- ----------
AA-IIa   541  TGCAACGAGT GAGATTGC-A A-----GTCT TTG--TT--- ----------
NA-IIb   539  TGCAACGAGT GAGATTGC-A A-----GCCT TTG--TT--- ----------
AA-IIc   537  TGCAACGAGT GAGATTGC-A A-----GCCT TTG--TT--- ----------
  CS-I   444  TGCAACGAGT GAGATTGC-G A-----GTCT TTG--TT--- ----------
 Align   240  mmmmmmmmmm mmmmmmmm-m m-----mmmm mmm--mm--- ----------

NA-IIa   569  ----ATGTAG TGAACAACAT ATACCTACTA CT-C------ -TAGTAGT--
AA-IIa   570  ----ATGTAG TGAACAACAT ATACCTACTA CT-C------ -TAGTAGT--
```

```
NA-IIb   568  ----ATGTAG TGAACAACAT ATACCTACTA CTACTACTAT TTTGTAGTAA
AA-IIc   566  ----ATGTAG TGAACAACAT ATACCTACTA CTACTACTT- -TGTAGTAAT
  CS-I   473  ----ATGTAG TGCTCAACAT A--CCTA--- ---------- -TGTGCAGT-
 Align   269  ----mmmmmm mmmmmmmmmm m--mmmm--- ---------- ----------

NA-IIa   605  --------AG GATTCCAACT ACACA---GT ATAACCATT- -----AAGGT
AA-IIa   606  --------AG GATTCCAACT ACACA---GT ATAACCATT- -----AAGGT
NA-IIb   614  TGT-----AG GATTCCAACT ACACA---GA ATAACCATT- -----AAGGT
AA-IIc   610  GT------AG GATTCCAACT ACACACA-GA ATAACCATT- -----AAGGT
  CS-I   502  --------AG GATTT-AACT ACACC---AT GTAACCTTTT TC--AAAGGT
 Align   290  --------mm mmmm--mmmm mmmm------ -mmmmmmmm- -----mmmmm

NA-IIa   638  GA-------- ---------- ---------- ---------- ----------
AA-IIa   639  GA-------- ---------- ---------- ---------- ----------
NA-IIb   650  GA-------- ---------- ---------- ---------- ----------
AA-IIc   647  GA-------- ---------- ---------- ---------- ----------
  CS-I   538  TG-------- ---------- ---------- ---------- ----------
 Align   317  ---------- ---------- ---------- ---------- ----------

NA-IIa   640  ---------- ---------- ---------A ATG-AACTCA G----GCGAC
AA-IIa   641  ---------- ---------- ---------A ATG-AACTCA G----GCGAC
NA-IIb   652  ---------- ---------- ---------A ATG-AACTCA G----GCGAC
AA-IIc   649  ---------- ---------- ---------A ATG-AACTCA G----GCGAC
  CS-I   540  ---------- ---------- ---------A ATG-AACTTA G----GCGAC
 Align   317  ---------- ---------- ---------m mmm-mmmmmm m----mmmmm

NA-IIa   656  TGCT------ ---------- ---------- --ATACTT-- -------T-A
AA-IIa   657  TGCT------ ---------- ---------- --ATACTT-- -------T-A
NA-IIb   668  TGCT------ ---------- ---------- --ATACTT-- -------T-A
AA-IIc   665  TGCT------ ---------- ---------- --ATACTT-- -------T-A
  CS-I   556  TGCT------ ---------- ---------- --ATACTTT- -------T-A
 Align   333  mmmm------ ---------- ---------- --mmmmmm-- -------m-m

NA-IIa   668  A-GATG---- ---------- ---------G TGGAAGGTTG TGGCAATGAC
AA-IIa   669  A-GATG---- ---------- ---------G TGGAAGGTTG TGGCAATGAC
NA-IIb   680  A-GATG---- ---------- ---------G TGGAAGGTTG TGGCAATGAC
AA-IIc   677  A-GATG---- ---------- ---------G TGGAAGGTTG TGGCAATGAC
  CS-I   569  A-GATG---- ---------- ---------G TCGAAGGTTG TGGCAATGAC
 Align   345  m-mmmm---- ---------- ---------m mmmmmmmmmmm mmmmmmmmmmm

NA-IIa   694  AGGTCTG-TG ATGCCCTTTA G-ATGTTCAA GG-CTGCACA CGTACTACAT
AA-IIa   695  AGGTCTG-TG ATGCCCTTTA G-ATGTTCAA GG-CTGCACA CGTACTACAT
NA-IIb   706  AGGTCTG-TG ATGCCCTTTA G-ATGTTCAA GG-CTGCACA CGTACTACAT
AA-IIc   703  AGGTCTG-TG ATGCCCTTTA G-ATGTTCAA GG-CTGCACA CGTACTACAT
  CS-I   595  AGGTCTG-TG ATGCCC-TTA G-ATGTTCAG GG-CTGCACA CGTACTACAT
 Align   371  mmmmmmmm-mm mmmmmm-mmm m-mmmmmmmmm mm-mmmmmmm mmmmmmmmmm

NA-IIa   741  TGA-TC-TAG T-CAACGAGT ATGTATGTAA CATTG----- ----------
AA-IIa   742  TGA-TC-TAG T-CAACGAGT ATGTATGTAA CATTG----- ----------
NA-IIb   753  TGA-TC-TAG T-CAACGAGT ATGTATGTAA CATTG----- ----------
AA-IIc   750  TGA-TC-TAG T-CAACGAGT ATGTATGTAA CATTG----- ----------
  CS-I   641  TGA-TC-TAG T-CAATAAGT ATGTGTGTAA C--------- ----------
 Align   417  mmm-mm-mmm m-mmmmmmmm mmmmmmmmmm m--------- ----------

NA-IIa   773  ----AATAAT GAATGTATTG GTTAAGCTAT ATTGTATTTC --GCTAA--C
AA-IIa   774  ----AATAAT GAATGTATTG GTTAAGCTAT ATTGTATTTC --GCTAA--C
NA-IIb   785  -----ATTTT GCATCTATTG GTTAAGCTAT AATGTATTTC --GGTAA--C
AA-IIc   782  -----ATTAT GCATCTATTG GTTAAGCTAT AATGTATTTC --GGTAA--C
  CS-I   669  -AATAATTTT GAATGTATTG GTTAAGCTTT TCTTATATTT T-GGTAAGAG
 Align   445  -----mmmmm mmmmmmmmmm mmmmmmmm-- ---------- --mmmmm---

NA-IIa   815  GTTAAT---- ---------- ---------- ---------- ----------
AA-IIa   816  GTTAAT---- ---------- ---------- ---------- ----------
NA-IIb   826  GTTAAT---- ---------- ---------- ---------- ----------
AA-IIc   823  GTTAAT---- ---------- ---------- ---------- ----------
  CS-I   717  GTTAAT---- ---------- ---------- ---------- ----------
 Align   473  mmmmmm---- ---------- ---------- ---------- ----------

NA-IIa   821  ---------- ---------- -----TCAGA ACTT--CG-A G-----AAGG
AA-IIa   822  ---------- ---------- -----TCAGA ACTT--CG-A G-----AAGG
NA-IIb   832  ---------- ---------- -----TCAGA ACTT--CG-A G-----AAGG
AA-IIc   829  ---------- ---------- -----TCAGA ACTT--CG-A G-----AAGG
  CS-I   723  ---------- ---------- -----ACAGA ACTT--CG-A G-----AGAG
 Align   479  ---------- ---------- -----mmmmm mmmm--mm-m m-----mmmm

NA-IIa   838  ---------- -----TTCCG -ACAACAGTC AATCTAGTTA TT----GCTT
AA-IIa   839  ---------- -----TTCCG -ACAACAGTC AATCTAGTTA TT----GCTT
```

```
NA-IIb   849  ---------- -----TTCCG -ACAACAGTC AATCTAGTTA TT----GCTT
AA-IIc   846  ---------- -----TTCCG -ACAACAGTC AATCTAGTTA TT----GCTT
  CS-I   740  ---------- -----TTCCG -ACAACAGTC AATCTAGTTA TT----GCTT
 Align   496  ---------- -----mmmmm -mmmmmmmmmm mmmmmmmmmmmm mm----mmmm

NA-IIa   868  GTAGTCG--- -ATTGCATAA TTGTACAGTT -CACGCAACG CTCGGTTATC
AA-IIa   869  GTAGTCG--- -ATTGCATAA TTGTACAGTT -CACGCAACG CTCGGTTATC
NA-IIb   879  GTAGTTGTTA -ATTGTATAA TTGTCTAGTT -CACGCAACG TTCGTTTATC
AA-IIc   876  GTAGTTGTTA -ATTGTATAA TTGTCTAGTT -CACGCAACG TTCGTTTATC
  CS-I   770  GTAGTCG--- -ATCTCATGA TTGTTTAGTT TTATGCAACT TCGGGTCATG
 Align   526  mmmmmmmm--- -mmmmmmmmmm mmmmmmmmmmm ---------- ----------

NA-IIa   913  ATT------- ---------- ---------- ---------- ----------
AA-IIa   914  ATT------- ---------- ---------- ---------- ----------
NA-IIb   927  ATT------- ---------- ---------- ---------- ----------
AA-IIc   924  ATT------- ---------- ---------- ---------- ----------
  CS-I   816  ATTCA----- ---------- ---------- ---------- ----------
 Align   552  ---------- ---------- ---------- ---------- ----------

NA-IIa   916  CGCTA--GTA --TTGTTAAT T-----CACA --GTG-GGGA CAGTCGTTTG
AA-IIa   917  CGCTA--GTA --TTGTTAAT T-----CACA --GTG-GGGA CAGTCGTTTG
NA-IIb   930  CTCTA--GTA --TTGTTAAT T-----CACA --GTG-GGGA CAGTCGTTTG
AA-IIc   927  CTCTA--GTA --TTGTTAAT T-----CACA --GTG-GGGA CAGTCGTTTG
  CS-I   821  CTAGT--ACG --CTTCTAAT T-----CACA --GTG-GGGA CAGTCGTTTG
 Align   552  ---------- ------mmmm m-----mmmm --mmm-mmmm mmmmmmmmmm

NA-IIa   954  ---TAATT-C TGAGACTCGG T-TCAACCAG GAATGCCTCG TATTT-GTG-
AA-IIa   955  ---TAATT-C TGAGACTCGG T-TCAACCAG GAATGCCTCG TATTT-GTG-
NA-IIb   968  ---TAATT-C TGAGACTCGG T-TCAACCAG GAATGCCTCG TATTT-GTG-
AA-IIc   965  ---TAATT-C TGAGACTCGG T-TCAACCAG GAATGCCTCG TATTT-GTG-
  CS-I   859  ---TAATT-C TGAGACTCGG T-TCAACTAG GAATGCCTAG TATTG-ATG-
 Align   578  ---mmmmm-m mmmmmmmmmm m-mmmmmmmm mmmmmmmmmm mmmmm-mmm-

NA-IIa   997  ---GTTCATT -ATACTCCGT GGAATAAGTC CCTGTCCTTT GTACACACCG
AA-IIa   998  ---GTTCATT -ATACTCCGT GGAATAAGTC CCTGTCCTTT GTACACACCG
NA-IIb  1011  ---GTTCATT -AAACTCCGT GGAATAAGTC CCTGTCCTTT GTACACACCG
AA-IIc  1008  ---GTTCATT -AAACTCCGT GGAATAAGTC CCTGTCCTTT GTACACACCG
  CS-I   902  ---GTTCACT -AAACTTTCT GGAATAAGTC CCTGCCCTTT GTACACACCG
 Align   621  ---mmmmmmm -mmmmmmmmm mmmmmmmmmm mmmmmmmmmm mmmmmmmmm

NA-IIa  1043  CCCGTCGCTT T-TACCAA-- TG--GCCCTC G-TTG-TGAG ATA-GCTGGA
AA-IIa  1044  CCCGTCGCTT T-TACCAA-- TG--GCCCTC G-TTG-TGAG ATA-GCTGGA
NA-IIb  1057  CCCGTCGCTT T-TACCAA-- TG--GCCCTC G-TTG-TGAG ATA-GCTGGA
AA-IIc  1054  CCCGTCGCTT T-TACCAA-- TG--GCCCTC G-TTG-TGAG ATA-GCTGGA
  CS-I   948  CCCGTCGCTT T-TACCAA-- TG--GCCCTC G-TTG-TGAG TGA-GCTGGA
 Align   667  mmmmmmmmmm m-mmmmmm-- mm--mmmmmm m-mmm-mmmm mmm-mmmmmm

NA-IIa  1085  CAAG-----T ATTTA----- ---------- ---------- ----------
AA-IIa  1086  CAAG-----T ATTTA----- ---------- ---------- ----------
NA-IIb  1099  CAAG-----T ATTTA----- ---------- ---------- ----------
AA-IIc  1096  CAAG-----T ATTTA----- ---------- ---------- ----------
  CS-I   990  CAAG----TT ATTT------ ---------- ---------- ----------
 Align   709  mmmm------ ---------- ---------- ---------- ----------

NA-IIa  1095  ---------- ---------- ---------- ---------- ------CTAC
AA-IIa  1096  ---------- ---------- ---------- ---------- ------CTAC
NA-IIb  1109  ---------- ---------- ---------- ---------- ------CTAC
AA-IIc  1106  ---------- ---------- ---------- ---------- ------CTAC
  CS-I  1000  ---------- ---------- ---------- ---------- ------TAAC
 Align   713  ---------- ---------- ---------- ---------- --------mm

NA-IIa  1099  TTCAAAAGT- --TC---TCA AT-CAAGG-T TTGCTAA-AG GAAAA-AGAA
AA-IIa  1100  TTCAAAAGT- --TC---TCA AT-CAAGG-T TTGCTAA-AG GAAAA-AGAA
NA-IIb  1113  TTCAAAAGT- --TC---TCA AT-CAAGG-T TTGCTAA-AG GAAAA-AGAA
AA-IIc  1110  TTCAAAAGT- --TC---TCA AT-CAAGG-T TTGCTAA-AG GAAAA-AGAA
  CS-I  1004  TTGAAAAGT- --TC---TCA AT-CAAGG-T TTGCTAA-AG GAAAA-AGAA
 Align   715  mmmmmmmmm- --mm---mmm mm-mmmmm-m mmmmmmmm-mm mmmmm-mmmm

NA-IIa  1139  -G-TCGTAAC
AA-IIa  1140  -G-TCGTAAC
NA-IIb  1153  -G-TCGTAAC
AA-IIc  1150  -G-TCGTAAC
  CS-I  1044  -G-TCGTAAC
 Align   755  -m-mmmmmmm
```

A2.1.4. *Globigerinella* sp. 767 bp molecular phylogeny

```
   NA-SI      1  GCA--CCACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
    C-SI      1  GCA--CCACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
  C-SI(P)     1  GCA--CCACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
   NA-cal     1  GCA--CCACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
   C-SIIa     1  GCA--CTACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
  NA-SIIa     1  GCA--CCACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
  CS-SIIa     1  GCA--CCACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
  CA-SIIa     1  GCA--CCACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
  CA-SIIb     1  GCA--CCACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
  NA-SIIb     1  GCA--CCACA AG--AGCGT- --GGAGCA-T GTGGCTTAAT TTGACTCAAC
    Align     1  mmm--mmmmm mm--mmmmm- --mmmmmm-m mmmmmmmmmmm mmmmmmmmmmm

   NA-SI     43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
    C-SI     43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
  C-SI(P)    43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
   NA-cal    43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAT-ACTGA G--GATTG-A
   C-SIIa    43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
  NA-SIIa    43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
  CS-SIIa    43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
  CA-SIIa    43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
  CA-SIIb    43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
  NA-SIIb    43  GCGGG-GAAT CTTACCGGG- -----TCCGG ACAC-ACTGA G--GATTG-A
    Align    43  mmmmm-mmmm mmmmmmmmmm- -----mmmmm mmmm-mmmmm m--mmmmm-m

   NA-SI     82  CAGACAGTT- -----GTCTT T-----CCCT CCC------- ----------
    C-SI     82  CAGACAGTT- -----GTCTT T-----CCCT CCC------- ----------
  C-SI(P)    82  CAGACAGTT- -----GTCTT T-----CCCT CCC------- ----------
   NA-cal    82  CAGACATCA- -CCACGTCTT TTTGTCCTTA AAA------- ----------
   C-SIIa    82  CAGACAATT- -----GTCTT T--TGGTCAA ATTTAA---- ----------
  NA-SIIa    82  CAGACAATT- -----GTCTT T--TTGTCAA ATTTAA---- ----------
  CS-SIIa    82  CAGACAATT- -----GTCTT T--TTGTCAA ATTAAA---- ----------
  CA-SIIa    82  CAGACAATT- -----GTCTT T--TTGTCAA ATTTAA---- ----------
  CA-SIIb    82  CAGACAATT- -----GTCCT T--TTGGCTT AAA------- ----------
  NA-SIIb    82  CAGACAATT- -----GTCCT T--TTGGCTT AAA------- ----------
    Align    82  mmmmmmmmm- -----mmmmm m--------- ---------- ----------

   NA-SI    104  ---------- ---TTC---G GGGTGGGA-- ---------- ----------
    C-SI    104  ---------- ---TTC---G GGGTGGGA-- ---------- ----------
  C-SI(P)   104  ---------- ---TTC---G GGGTGGGA-- ---------- ----------
   NA-cal   113  ---------- ---------- ---------- ---------- ----------
   C-SIIa   110  ---------- ---------- ---------- ---------- ----------
  NA-SIIa   110  ---------- ---------- ---------- ---------- ----------
  CS-SIIa   110  ---------- ---------- ---------- ---------- ----------
  CA-SIIa   110  ---------- ---------- ---------- ---------- ----------
  CA-SIIb   107  ---------- ---------- ---------- ---------- ----------
  NA-SIIb   107  ---------- ---------- ---------- ---------- ----------
    Align    97  ---------- ---------- ---------- ---------- ----------

   NA-SI    116  ---------- ---------- ---------- ------TTAC AAAAGATAAC
    C-SI    116  ---------- ---------- ---------- ------TTAC AAAAGATAAC
  C-SI(P)   116  ---------- ---------- ---------- ------TTAC AAAAGATAAC
   NA-cal   113  ---------- ---------- ---------- ------AGAC AAAAGATAAC
   C-SIIa   110  ---------- ---------- ---------- ------AGAC AAAAGATAAT
  NA-SIIa   110  ---------- ---------- ---------- ------AGAC AAAAGATAAC
  CS-SIIa   110  ---------- ---------- ---------- ------AGAC AAAAGATAAC
  CA-SIIa   110  ---------- ---------- ---------- ------AGAC AAAAGATAAC
  CA-SIIb   107  ---------- ---------- ---------- ------AGAC AAAAGATAAC
  NA-SIIb   107  ---------- ---------- ---------- ------AGAC AAAAGATAAC
    Align    97  ---------- ---------- ---------- ------mmmm mmmmmmmmmmm

   NA-SI    130  A-GA------ ---------- ---TCTTTCA TGA------T CATGTGATGG
    C-SI    130  A-GA------ ---------- ---TCTTTCA TGA------T CATGTGATGG
  C-SI(P)   130  A-GA------ ---------- ---TCTTTCA TGA------T CATGTGATGG
   NA-cal   127  A-GA------ ---------- ---TCTTTCA TGA------T TATGTGGTGG
   C-SIIa   124  A-GA------ ---------- ---TCTTTCA TGA------T TATGTGATGG
  NA-SIIa   124  A-GA------ ---------- ---TCTTTCA TGA------T TATGTGATGG
  CS-SIIa   124  A-GA------ ---------- ---TCTTTCA TGA------T TATGTGATGG
  CA-SIIa   124  A-GA------ ---------- ---TCTTTCA TGA------T TATGTGATGG
  CA-SIIb   121  A-GA------ ---------- ---TCTTTCA TGA------T TATGTGATGG
  NA-SIIb   121  A-GA------ ---------- ---TCTTTCA TGA------T TATGTGATGG
    Align   111  m-mm------ ---------- ---mmmmmmm mmm------m mmmmmmmmmmm
```

```
  NA-SI   154  -----GTGGT G-CATGG-CC GTT-CTTAGT TCGTGGA-GT GATC-TGTC-
   C-SI   154  -----GTGGT G-CATGG-CC GTT-CTTAGT TCGTGGA-GT GATC-TGTC-
 C-SI(P)  154  -----GTGGT G-CATGG-CC GTT-CTTAGT TCGTGGA-GT GATC-TGTC-
  NA-cal  151  -----GTGGT G-CATGG-CC GTT-CTTAGT TCGTGGA-GT GATC-TGTC-
  C-SIIa  148  -----GTGGT G-CATGG-CC GTT-CTTAGT TCGTGGA-GT GATC-TGTC-
 NA-SIIa  148  -----GTGGT G-CATGG-CC GTT-CTTAGT TCGTGGA-GT GATC-TGTC-
 CS-SIIa  148  -----GTGGT G-CATGG-CC GTT-CTTAGT TCGTGGA-GT GATC-TGTC-
 CA-SIIa  148  -----GTGGT G-CATGG-CC GTT-CTTAGT TCGTGGA-GT GATC-TGTC-
 CA-SIIb  145  -----GTGGT G-CATGG-CC GTT-CTTAGT TCGTGGA-GT GATC-TGTC-
 NA-SIIb  145  -----GTGGT G-CATGG-CC GTT-CTTAGT TCGTGGA-GT GATC-TGTC-
   Align  135  -----mmmmm m-mmmmm-mm mmm-mmmmmm mmmmmmm-mm mmmm-mmmm-

  NA-SI   193  -TGCTT-AAT TGCGTTTC-- -----AAATT TTATATTTGT TAACATGC-G
   C-SI   193  -TGCTT-AAT TGCGTTTC-- -----AAATT TTATATTTGT TAACATGC-G
 C-SI(P)  193  -TGCTT-AAT TGCGTTTC-- -----AAATT TTATATTTGT TAACATGC-G
  NA-cal  190  -TGCTT-AAT TGCGTTTC-- -----AAATG TATTTTTGAT TTCT------
  C-SIIa  187  -TGCTT-AAT TGCGTTTC-- -----AAATA TATTTTATTA TATAAC---G
 NA-SIIa  187  -TGCTT-AAT TGCGTTTC-- -----AAATA TATTTTATTA TATAAC---G
 CS-SIIa  187  -TGCTT-AAT TGCGTTTC-- -----AAATA TATTTTATTA TATA-C---G
 CA-SIIa  187  -TGCTT-AAT TGCGTTTC-- -----AAATA TATTTTATTA TATAAC---G
 CA-SIIb  184  -TGCTT-AAT TGCGTTTC-- -----AAATA TACTTTTTTT ACAA-C---G
 NA-SIIb  184  -TGCTT-AAT TGCGTTTC-- -----AAATA TACTTTTTTT ACAA-C---G
   Align  174  -mmmmm-mmm mmmmmmmmm-- -----mmmmm mm-------- ----------

  NA-SI   233  AGTCTTCAAT -----GACGG CCGCTCCACC TTATAGACAT TATATGATAT
   C-SI   233  AGTCTTCAAT -----GACGG CCGCTCCACC TTATAGACAT TATATGATAT
 C-SI(P)  233  AGTCTTCAAT -----GACGG CCGCTCCACC TTATAGACAT TATATGATAT
  NA-cal  225  ------CAAA -----CGAGT CTACGAAGAT GTCTATTCAT AACAAGCACT
  C-SIIa  225  AGTCTACAAA -----GGAGT GGTTCATTAA CAGACAATCA TATCTTTGAG
 NA-SIIa  225  AGTCTACAAA -----GGAGT GGTTCATTAA CAGACAATCA TATCTTTGAG
 CS-SIIa  224  AGTCTACAAA -----GGAGT GGTTCATTAA CAGACAATCA TATCTTTGAG
 CA-SIIa  225  AGTCTACAAA -----GGAGT GGTTCATTAA CAGACAATCA TATCTTTGAG
 CA-SIIb  221  AGTCTACAAA -----GGAGT GGTTCAGTTA CAACAGACAA TTAATTCTTG
 NA-SIIb  221  AGTCTACAAA -----GGAGT GGTTCAGTTA CAACAGACAA TTAATTCTTG
   Align  197  ------mmmm ---------- ---------- ---------- ----------

  NA-SI   278  GCACATTTGT GTATTTGATT ATAACTTGTC TGGA--GTCT GG-CTCGATT
   C-SI   278  GCACATTTGT GTATTTGATT ATAACTTGTC TGGA--GTCT GG-CTCGATT
 C-SI(P)  278  GCACATTTGT GTATTTGATT ATAACTTGTC TGGA--GTCT GG-CTCGATT
  NA-cal  264  TCATTTATGA TTTGTTTGGA ATACGACTCT TTC------- ----------
  C-SIIa  270  CGTCTGG-AA TCTACTCTAT TT-------- ---------- ----------
 NA-SIIa  270  CGTCTGG-AA TCTACTCTAT TT-------- ---------- ----------
 CS-SIIa  269  CGTCTGG-AA TCTACTCTAT TT-------- ---------- ----------
 CA-SIIa  270  CGTCTGG-AA TCTACTCTAT TT-------- ---------- ----------
 CA-SIIb  266  TGTTTTG-AA CGTTTGGAAT GTACTCTATT T--------- ----------
 NA-SIIb  266  TGTTTTG-AA CGTTTGGAAT GTACTCTATT T--------- ----------
   Align  201  ---------- ---------- ---------- ---------- ----------

  NA-SI   325  TTTTT----- ---------- ---------- ---------- ----------
   C-SI   325  TTTTT----- ---------- ---------- ---------- ----------
 C-SI(P)  325  TTTTT----- ---------- ---------- ---------- ----------
  NA-cal  297  ---------- ---------- ---------- ---------- ----------
  C-SIIa  291  ---------- ---------- ---------- ---------- ----------
 NA-SIIa  291  ---------- ---------- ---------- ---------- ----------
 CS-SIIa  290  ---------- ---------- ---------- ---------- ----------
 CA-SIIa  291  ---------- ---------- ---------- ---------- ----------
 CA-SIIb  296  ---------- ---------- ---------- ---------- ----------
 NA-SIIb  296  ---------- ---------- ---------- ---------- ----------
   Align  201  ---------- ---------- ---------- ---------- ----------


  NA-SI   330  ---------- ---------- ---------- ---GACTCAA TTGAACGCAA
   C-SI   330  ---------- ---------- ---------- ---GACTCAA TTGAACGCAA
 C-SI(P)  330  ---------- ---------- ---------- ---GACTCAA TTGAACGCAA
  NA-cal  297  ---------- ---------- ---------- ---GGCTCAA TTGAACGCAA
  C-SIIa  291  ---------- ---------- ---------- ---GGCTCAA TTGAACGCAA
 NA-SIIa  291  ---------- ---------- ---------- ---GGCTCAA TTGAACGCAA
 CS-SIIa  290  ---------- ---------- ---------- ---GGCTCAA TTGAACGCAA
 CA-SIIa  291  ---------- ---------- ---------- ---GGCTCAA TTGAACGCAA
 CA-SIIb  296  ---------- ---------- ---------- ---GGCTCAA TTGAACGCAA
 NA-SIIb  296  ---------- ---------- ---------- ---GGCTCAA TTGAACGCAA
   Align  201  ---------- ---------- ---------- ---mmmmmmm mmmmmmmmmmm

  NA-SI   347  CGGACGTGAT TGC-AA---- GTCCTTGT-- TG-------- ---------A
   C-SI   347  CGGACGTGAT TGC-AA---- GTCCTTGT-- TG-------- ---------A
 C-SI(P)  347  CGGACGTGAT TGC-AA---- GTCCTTGT-- TG-------- ---------A
  NA-cal  314  CGGACGTGAT TGC-AA---- GTCCTTGT-- TG-------- ---------A
```

```
 C-SIIa   308  CGGACGTGAT CGC-GA---- GTCCTTGT-- TG-------- ---------A
NA-SIIa   308  CGGACGTGAT CGC-GA---- GTCCTTGT-- TG-------- ---------A
CS-SIIa   307  CGGACGTGAT CGC-GA---- GTCCTTGT-- TG-------- ---------A
CA-SIIa   308  CGGACGTGAT CGC-GA---- GTCCTTGT-- TG-------- ---------A
CA-SIIb   313  CGGACGTGAT CGC-AA---- GTCCTTGT-- TG-------- ---------A
NA-SIIb   313  CGGACGTGAT CGC-AA---- GTCCTTGT-- TG-------- ---------A
  Align   218  mmmmmmmmmmmm mmm-mm---- mmmmmmmmmm-- mm-------- ---------m

  NA-SI   373  ACA-AATT-A -TATATAT-- A-CTACTTCA TCATTA---- GTATATATA-
   C-SI   373  ACA-AATT-A -TATATAT-- A-CTACTTCA TCATTA---- GTATATATA-
C-SI(P)   373  ACA-AATT-A -TATATAT-- A-CTACTTCA TCATTA---- GTATATATA-
 NA-cal   340  ACT-TC---- -TATACAT-A TCATTTACAT TCTTAACTGA GTGTATCTAA
 C-SIIa   334  ACT-TCAG-- -TATATAT-T GACTATCCCC GTAGGAAGGG ATTGGCAATA
NA-SIIa   334  ACT-TCAG-- -TATATAT-T GACTATCCCC GTAGGAAGGG ATTGGCAATA
CS-SIIa   333  ACT-TCAG-- -TATATAT-T GACTTTCCCC GTAGGAAGGG ATTGGCAATA
CA-SIIa   334  ACT-TCAG-- -TATATAT-T GACTATCCCC GTAGGAAGGG ATTGGCAATA
CA-SIIb   339  ACT-TCAG-- -TATATAT-T GACTTTCCCC GTAGGAAGGG ATTGCCAATA
NA-SIIb   339  ACT-TCAG-- -TATATAT-T GACTTTCCCC GTAGGAAGGG ATTGCCAATA
  Align   244  mmm-------- -mmmmmmmm-- ---------- ---------- ----------

  NA-SI   412  ---------G TTCGCT-TCT CATG-TCGTA G--CAGTCAA ACGGGCGGGC
   C-SI   412  ---------G TTCGCT-TCT CATG-TCGTA G--CAGTCAA ACGGGCGGGC
C-SI(P)   412  ---------G TTCGCT-TCT CATG-TCGTA G--CAGTCAA ACGGGCGGGC
 NA-cal   383  GATAATAGAG TTCGCTTTCT -ATT-GGTGA CGGGATCTCC AATCACTTTC
 C-SIIa   379  ATTGAAT--G TTCGCTTTCT -ATTAGGTGT GTATATATAG TCTCCAAATT
NA-SIIa   379  ATTGAAT--G TTCGCTTTCT -ATTAGGTGT GTATATATAG TCTCAAAATT
CS-SIIa   378  ATTGAAT--G TTCGCTTTCT -ATTAGGTTT GTATGT--AA TCTCAAAATT
CA-SIIa   379  ATTGAAT--G TTCGCTTTCT -ATTAGGTGT GTATATATAG TCTCAAAATT
CA-SIIb   384  ATTGAAT--G TTCGCTTTCT -ATT-GGTGT GTGATCTCAA AATTGCATGT
NA-SIIb   384  ATTGAAT--G TTCGCTTTCT -ATT-GGTGT GTGATCTCAA AATTGCATGT
  Align   254  ---------m mmmmmm-mmm -mmm------ ---------- ----------

  NA-SI   449  GTCTTAATTG GCGCGGTCAC GAGG---CTT AATATGCTGC GGCAG-----
   C-SI   449  GTCTTAATTG GCGCGGTCAC GAGG---CTT AATATGCTGC GGCAG-----
C-SI(P)   449  GTCTTAATTG GCGCGGTCAC GAGG---CTT AATATGCTGC GGCAG-----
 NA-cal   431  GAGTGAGAGT --GAGTGATC TAATCCTTTA ---------- ----------
 C-SIIa   426  ACATGTATTT GTAATGGCGA GTAATATTTA --CTATCCTG TA--------
NA-SIIa   426  ACATGTATTT GTAATGGCGA GTAATATTTA TACTATCCTG TA--------
CS-SIIa   423  ACATGTATTT GTAATGGCGA GTGAT---CA --CA--CCTA TA--------
CA-SIIa   426  ACATGTATTT GTAATGGCGA GTAATATATT --ACTATCCT GTA-------
CA-SIIb   430  ATTTGTGATG GCGAGTGATC TAATCCCTTA ---------- ----------
NA-SIIb   430  ATTTGTGATG GCGAGTGATC TAATCCCTTA ---------- ----------
  Align   267  ---------- ---------- ---------- ---------- ----------


  NA-SI   491  ---------- ---------- ---------- ---------- --------GG
   C-SI   491  ---------- ---------- ---------- ---------- --------GG
C-SI(P)   491  ---------- ---------- ---------- ---------- --------GG
 NA-cal   459  ---------- ---------- ---------- ---------- --------GG
 C-SIIa   466  ---------- ---------- ---------- ---------- --------GG
NA-SIIa   468  ---------- ---------- ---------- ---------- --------GG
CS-SIIa   458  ---------- ---------- ---------- ---------- --------GG
CA-SIIa   467  ---------- ---------- ---------- ---------- --------GG
CA-SIIb   460  ---------- ---------- ---------- ---------- --------GG
NA-SIIb   460  ---------- ---------- ---------- ---------- --------GG
  Align   267  ---------- ---------- ---------- ---------- --------mm

  NA-SI   493  AA-AACTTGG ----GCGACC GCT------- ---------- -----GTAAT
   C-SI   493  AA-AACTTGG ----GCGACC GCT------- ---------- -----GTAAT
C-SI(P)   493  AA-AACTTGG ----GCGACC GCT------- ---------- -----GTAAT
 NA-cal   461  AA-AACTCGG ----GCGACC GCT------- ---------- -----GTAAT
 C-SIIa   468  AA-AACTCGG ----GCGACC GCT------- ---------- -----GTAAT
NA-SIIa   470  AA-AACTCGG ----GCGACC GCT------- ---------- -----GTAAT
CS-SIIa   460  AA-AACTCGG ----GCGACC GCT------- ---------- -----GTAAT
CA-SIIa   469  AA-AACTCGG ----GCGACC GCT------- ---------- -----GTAAT
CA-SIIb   462  AA-AACTTGG ----GCGACC GCT------- ---------- -----GTAAT
NA-SIIb   462  AA-AACTTGG ----GCGACC GCT------- ---------- -----GTAAT
  Align   269  mm-mmmmmmmm ----mmmmmm mmm------- ---------- -----mmmmm

  NA-SI   516  ACTTCTCTCT ------T-AA -ACCA----- ---------- --------GA
   C-SI   516  ACTTCTCTCT ------T-AA -ACCA----- ---------- --------GA
C-SI(P)   516  ACTTCTCTCT ------T-AA -ACCA----- ---------- --------GA
 NA-cal   484  ACTTTCTTTT ------T-AA -AACA----- ---------- --------GA
 C-SIIa   491  ATTTCTCT-T ------T-AA -AACA----- ---------- --------GA
NA-SIIa   493  ATTTCTCT-T ------T-AA -AACA----- ---------- --------GA
CS-SIIa   483  ATTTCTCT-T ------T-AA -AACA----- ---------- --------GA
CA-SIIa   492  ATTTCTCT-T ------T-AA -AACA----- ---------- --------GA
```

```
CA-SIIb   485  ACTTCTCT-T ------T-AA -AACA----- ---------- --------GA
NA-SIIb   485  ACTTCTCT-T ------T-AA -AACA----- ---------- --------GA
  Align   292  mmmmmmmmm-m ------m-mm -mmmm----- ---------- --------mm

  NA-SI   535  GGAAGATTGC GGCAATAACA GGTCTG-TGA TGCCC-TCAG -ATGTCCCGG
   C-SI   535  GGAAGATTGC GGCAATAACA GGTCTG-TGA TGCCC-TCAG -ATGTCCCGG
C-SI(P)   535  GGAAGATTGC GGCAATAACA GGTCTG-TGA TGCCC-TCAG -ATGTCCCGG
 NA-cal   503  GGAAGATTGC GGCAATAACA GGTCTG-TGA TGCCC-TCAG -ATGTCCCGG
 C-SIIa   509  GGAAGATTGC GGCAATAACA GGTCTG-TGA TGCCC-TCAG -ATGTCCCGG
NA-SIIa   511  GGAAGATTGC GGCAATAACA GGTCTG-TGA TGCCC-TCAG -ATGTCCCGG
CS-SIIa   501  GGAAGATTGC GGCAATAACA GGTCTG-TGA TGCCC-TCAG -ATGTCCCGG
CA-SIIa   510  GGAAGATTGC GGCAATAACA GGTCTG-TGA TGCCC-TCAG -ATGTCCCGG
CA-SIIb   503  GGCAGGTTGC GGCAATAACA GGTCTG-TGA TGCCC-TCAG -ATGTCCCGG
NA-SIIb   503  GGCAGGTTGC GGCAATAACA GGTCTG-TGA TGCCC-TCAG -ATGTCCCGG
  Align   310  mmmmmmmmmmmm mmmmmmmmmmmm mmmmmmm-mmm mmmmm-mmmm -mmmmmmmmmm

  NA-SI   582  G-CCGCACAC GTGCTACATT GA-TT-AGCG -CAGTGCGCA TACTTACA-T
   C-SI   582  G-CCGCACAC GTGCTACATT GA-TT-AGCG -CAGTGCGCA TACTTACA-T
C-SI(P)   582  G-CCGCACAC GTGCTACATT GA-TT-AGCG -CAGTGCGCA TACTTACA-T
 NA-cal   550  G-CCGCACAC GTGCTACATT GA-TT-AGCG -CAGTGCGCA TACTTTCAAT
 C-SIIa   556  G-CCGCACAC GTGCTACATT GA-TT-AGCG -CAGTGCGCA TACTTTCAAT
NA-SIIa   558  G-CCGCACAC GTGCTACATT GA-TT-AGCG -CAGTGCGCA TACTTTCAAT
CS-SIIa   548  G-CCGCACAC GTGCTACATT GA-TT-AGCG -CAGTGCGCA TACTTTCAAT
CA-SIIa   557  G-CCGCACAC GTGCTACATT GA-TT-AGCG -CAGTGCGCA TACTTTCAAT
CA-SIIb   550  G-CCGCACAC GTGCTACATT GA-TT-AGCG -CAGTGCGCA TACTTTCAAT
NA-SIIb   550  G-CCGCACAC GTGCTACATT GA-TT-AGCG -CAGTGCGCA TACTTTCAAT
  Align   357  m-mmmmmmmmm mmmmmmmmmmmm mm-mm-mmmm -mmmmmmmmmm mmmmmmmmm-m


  NA-SI   627  AT-------- ---------- ---AG-ATCA TT----G-AT T--G-G-TTA
   C-SI   627  AT-------- ---------- ---AG-ATCA TT----G-AT T--G-G-TTA
C-SI(P)   627  AT-------- ---------- ---AG-ATCA TT----G-AT T--G-G-TTA
 NA-cal   596  AT-------- ---------- ---GA-TACA TT---GG-TT TT-G-GTAGT
 C-SIIa   602  AT-------- ---------- ---GA-TACA TT---AG-AT T--G-GATTG
NA-SIIa   604  AT-------- ---------- ---GA-TACA TT---AG-AT T--G-GATTG
CS-SIIa   594  AT-------- ---------- ---GA-TACA TT---AG-AT T--G-GATAG
CA-SIIa   603  AT-------- ---------- ---GA-TACA TT---AG-AT T--G-GATTG
CA-SIIb   596  AT-------- ---------- ---GA-TACA TT---GG-AT T--G-GTGGT
NA-SIIb   596  AT-------- ---------- ---GA-TACA TT---GG-AT T--G-GTGGT
  Align   402  mm-------- ---------- ---mm-mmmm mm----m-mm m--m-m----

  NA-SI   646  TTT---ATAA CCG------- ---------- ---------- ----------
   C-SI   646  TTT---ATAA CCG------- ---------- ---------- ----------
C-SI(P)   646  TTT---ATAA CCG------- ---------- ---------- ----------
 NA-cal   618  AAGCTATTCA TCTGATGGAT ATGCTACTCT CCATACCAAT ACT-------
 C-SIIa   623  TTCTG----- -----TTGAG CT-CCAT--- ---------- ----------
NA-SIIa   625  TTCTG----- -----TTGAG CT-CCAT--- ---------- ----------
CS-SIIa   615  CTT------- -----TT--G CTTCCAT--- ---------- ----------
CA-SIIa   624  TTCTG----- -----TTGAG CT-CCAT--- ---------- ----------
CA-SIIb   617  GAATTGTTGG CCCCGTCTAA TACT------ ---------- ----------
NA-SIIb   617  GAATTGTTGG CCCCGTCTAA TACT------ ---------- ----------
  Align   418  ---------- ---------- ---------- ---------- ----------

  NA-SI   656  ----TCAATA TC-------- ---------- ---------- ----------
   C-SI   656  ----TCAATA TC-------- ---------- ---------- ----------
C-SI(P)   656  ----TCAATA TC-------- ---------- ---------- ----------
 NA-cal   661  ---------- ---------- ---------- ---------- ----------
 C-SIIa   639  ----CTAATA CT-------- ---------- ---------- ----------
NA-SIIa   641  ----CTAATA CT-------- ---------- ---------- ----------
CS-SIIa   628  ----CTAATA CT-------- ---------- ---------- ----------
CA-SIIa   640  ----CTAATA CT-------- ---------- ---------- ----------
CA-SIIb   641  ---------- ---------- ---------- ---------- ----------
NA-SIIb   641  ---------- ---------- ---------- ---------- ----------
  Align   418  ---------- ---------- ---------- ---------- ----------

  NA-SI   664  ---------- ---------- ---------- ---------- GCCTTGTCT-
   C-SI   664  ---------- ---------- ---------- ---------- GCCTTGTCT-
C-SI(P)   664  ---------- ---------- ---------- ---------- GCCTTGTCT-
 NA-cal   661  ---------- ---------- ---------- ---------- ATCTAGTCT-
 C-SIIa   647  ---------- ---------- ---------- ---------- ATCCGGCTT-
NA-SIIa   649  ---------- ---------- ---------- ---------- ATCCGGCTT-
CS-SIIa   636  ---------- ---------- ---------- ---------- ATCCGGCTT-
CA-SIIa   648  ---------- ---------- ---------- ---------- ATCCGGCTT-
CA-SIIb   641  ---------- ---------- ---------- ---------- GTCCTGTCT-
NA-SIIb   641  ---------- ---------- ---------- ---------- GTCCTGTCT-
  Align   418  ---------- ---------- ---------- ---------- mmmmmmmmmm-
```

```
  NA-SI   673  -GA-AA---- -GGAC----- ---------- -TAGG-TAAT CTATTGTAAG
   C-SI   673  -GA-AA---- -GGAC----- ---------- -TAGG-TAAT CTATTGTAAG
C-SI(P)   673  -GA-AA---- -GGAC----- ---------- -TAGG-TAAT CTATTGTAAG
  NA-cal  670  -GA-AA---- -AGAC----- ---------- -TGGG-TAAT CTATTGTAAG
  C-SIIa  656  -GA-GA---- -AGGC----- ---------- -TGGG-TAAT CAATTGTAAG
 NA-SIIa  658  -GA-GA---- -AGGC----- ---------- -TGGG-TAAT CAATTGTAAG
 CS-SIIa  645  -GA-GA---- -AGGC----- ---------- -TGGG-TAAT CAATTGTAAG
 CA-SIIa  657  -GA-GA---- -AGGC----- ---------- -TGGG-TAAT CAATTGTAAG
 CA-SIIb  650  -GA-GA---- -AGGC----- ---------- -TGGG-TAAT CAATTGTAAG
 NA-SIIb  650  -GA-GA---- -AGGC----- ---------- -TGGG-TAAT CAATTGTAAG
   Align  427  -mm-mm---- -mmmm----- ---------- -mmmm-mmmm mmmmmmmmmm


  NA-SI   699  TGCTGGT--- -TCC--TCCT ---CCCGTT- --G--AGCA- -TTTTA---A
   C-SI   699  TGCTGGT--- -TCC--TCCT ---CCCGTT- --G--AGCA- -TTTTA---A
C-SI(P)   699  TGCTGGT--- -TCC--TCCT ---CCCGTT- --G--AGCA- -TTTTA---A
  NA-cal  696  TGCTGGT--- -TCC--TCTT ---CCCGTT- --G--AGCA- -TTTTA---A
  C-SIIa  682  TGCTGGT--- -TCC--TCCT ---CCCGTT- --G--AGCA- -TTTTA---A
 NA-SIIa  684  TGCTGGT--- -TCC--TCCT ---CCCGTT- --G--AGCA- -TTTTA---A
 CS-SIIa  671  TGCTGGT--- -TCC--TCCT ---CCCGTT- --G--AGCA- -TTTTA---A
 CA-SIIa  683  TGCTGGT--- -TCC--TCCT ---CCCGTT- --G--AGCA- -TTTTA---A
 CA-SIIb  676  TGCTGGT--- -TCC--TACT ---CCCGTT- --G--AGCA- -TTTTA---A
 NA-SIIb  676  TGCTGGT--- -TCC--TACT ---CCCGTT- --G--AGCA- -TTTTA---A
   Align  453  mmmmmmmm--- -mmm--mmmm ---mmmmmm- --m--mmmm- -mmmmm---m

  NA-SI   730  TAATGG--TC --TCTCTACA TCCCTAGCAC ATGATG---T C--TAGTGCG
   C-SI   730  TAATGG--TC --TCTCTACA TCCCTAGCAC ATGATG---T C--TAGTGCG
C-SI(P)   730  TAATGG--TC --TCTCTACA TCCCTAGCAC ATGATG---T C--TAGTGCG
  NA-cal  727  TAATGG--TC --TCTCTACG TCCCTAGTAT AT-------T C--TAGTGCG
  C-SIIa  713  TAATGG--CC --TCTCTACA TCCCTAGCA- AT-ATG---C C--TAGTGCG
 NA-SIIa  715  TAATGG--CC --TCTCTACA TCCCTAGCA- AT-ATG---C C--TAGTGCG
 CS-SIIa  702  TAATGG--CC --TCTCTACA TCCCTAGTA- AT-ATG---A C--TAGTGCG
 CA-SIIa  714  TAATGG--CC --TCTCTACA TCCCTAGCA- AT-ATG---C C--TAGTGCG
 CA-SIIb  707  TAATGG--CC --TCTCTACA TCCCTAGTA- AC-------C C--TAGTGCG
 NA-SIIb  707  TAATGG--CC --TCTCTACA TCCCTAGTA- AC-------C C--TAGTGCG
   Align  484  mmmmmm--mm --mmmmmmmm mmmmmmmmm- mm-------m m--mmmmmm

  NA-SI   771  ATTG---TAG TTGAGTCTTG --CCATT--- ---------- ----------
   C-SI   771  ATTG---TAG TTGAGTCTTG --CCATT--- ---------- ----------
C-SI(P)   771  ATTG---TAG TTGAGTCTTG --CCATT--- ---------- ----------
  NA-cal  764  ATTG---TAG TTGAGTCT-G --CCATT--- ---------- ----------
  C-SIIa  752  ATTG---TAG TTGAGCCTTG --CCATT--- ---------- ----------
 NA-SIIa  754  ATTG---TAG TTGAGCCTTG --CCATT--- ---------- ----------
 CS-SIIa  741  ATTG---TAG TTGAGCCTTG --CCATT--- ---------- ----------
 CA-SIIa  753  ATTG---TAG TTGAGCCTTG --CCATT--- ---------- ----------
 CA-SIIb  743  ATTG---TAG TTGAGCCTTG --CCATT--- ---------- ----------
 NA-SIIb  743  ATTG---TAG TTGAGCCTTG --CCATT--- ---------- ----------
   Align  520  mmmm---mmm mmmmmmmmm-m --mmmmm--- ---------- ----------

  NA-SI   793  ---------- TATGC--AA- --GTGTCAAT T-----CTCA --GTG-GGGA
   C-SI   793  ---------- TATGC--AA- --GTGTCAAT T-----CTCA --GTG-GGGA
C-SI(P)   793  ---------- TATGC--AA- --GTGTCAAT T-----CTCA --GTG-GGGA
  NA-cal  785  ---------- TATGC--AA- --GTGTCAAT T-----CTCA --GTG-GGGA
  C-SIIa  774  ---------- TATGC--AA- --GTGTCAAT T-----CTCA --GTG-GGGA
 NA-SIIa  776  ---------- TATGC--AA- --GTGTCAAT T-----CTCA --GTG-GGGA
 CS-SIIa  763  ---------- TATGC--AA- --GTGTCAAT T-----CTCA --GTG-GGGA
 CA-SIIa  775  ---------- TATGC--AA- --GTGTCAAT T-----CTCA --GTG-GGGA
 CA-SIIb  765  ---------- TATGC--AA- --GTGTCAAT T-----CTCA --GTG-GGGA
 NA-SIIb  765  ---------- TATGC--AA- --GTGTCAAT T-----CTCA --GTG-GGGA
   Align  541  ---------- mmmmm--mm- --mmmmmmmm m-----mmmm --mmm-mmmm

  NA-SI   820  CAGCCATTTG A--TAATT-C TTTGGCTCGG CCTCAACTAG GAATGCCTTG
   C-SI   820  CAGCCATTTG A--TAATT-C TTTGGCTCGG CCTCAACTAG GAATGCCTTG
C-SI(P)   820  CAGCCATTTG A--TAATT-C TTTGGCTCGG CCTCAACTAG GAATGCCTTG
  NA-cal  812  CAGACATTTG A--TAATT-C TTTGTCTCGT TCTTAACTAG GAATGCCTTG
  C-SIIa  801  CAGACATTTG A--TAATT-C TTTGTCTCGT TCTTAACTAG GAATGCCTTG
 NA-SIIa  803  CAGACATTTG A--TAATT-C TTTGTCTCGT TCTTAACTAG GAATGCCTTG
 CS-SIIa  790  CAGACATTTG A--TAATT-C TTTGTCTCGT TCTTAACTAG GAATGCCTTG
 CA-SIIa  802  CAGACATTTG A--TAATT-C TTTGTCTCGT TCTTAACTAG GAATGCCTTG
 CA-SIIb  792  CAGACATTTG A--TAATT-C TTTGTCTCGT TCTTAACTAG GAATGCCTTG
 NA-SIIb  792  CAGACATTTG A--TAATT-C TTTGTCTCGT TCTTAACTAG GAATGCCTTG
   Align  568  mmmmmmmmmm m--mmmmm-m mmmmmmmmmm mmmmmmmmmm mmmmmmmmmm


  NA-SI   867  TACGG-GTCT T-GGTTCAAC -AGACCACCC GGAATACGTC CCTGCCCTTT
   C-SI   867  TACGG-GTCT T-GGTTCAAC -AGACCACCC GGAATACGTC CCTGCCCTTT
C-SI(P)   867  TACGG-GTCT T-GGTTCAAC -AGACCACCC GGAATACGTC CCTGCCCTTT
```

```
 NA-cal   859  TACGG-GTCT T-GGTTCAAC -AGACCACCC GGAATACGTC CCTGCCCTTT
 C-SIIa   848  TACGG-GTCT T-GGTTCAAC -AGACCACCC GGAATACGTC CCTGCCCTTT
NA-SIIa   850  TACGG-GTCT T-GGTTCAAC -AGACCACCC GGAATACGTC CCTGCCCTTT
CS-SIIa   837  TACGG-GTCT T-GGTTCAAC -AGACCACCC GGAATACGTC CCTGCCCTTT
CA-SIIa   849  TACGG-GTCT T-GGTTCAAC -AGACCACCC GGAATACGTC CCTGCCCTTT
CA-SIIb   839  TACGG-GTCT T-GGTTCAAC -AGACCACCC GGAATACGTC CCTGCCCTTT
NA-SIIb   839  TACGG-GTCT T-GGTTCAAC -AGACCACCC GGAATACGTC CCTGCCCTTT
  Align   615  mmmmm-mmmm m-immmmmmm -mmmmmmmmm mmmmmmmmmmm mmmmmmmmmm

  NA-SI   914  GTACACACCG CCCGTCGCTC T-TACCGA-- TG--AATTGT A-CTG-TGAG
   C-SI   914  GTACACACCG CCCGTCGCTC T-TACCGA-- TG--AATTGT A-CTG-TGAG
C-SI(P)   914  GTACACACCG CCCGTCGCTC T-TACCGA-- TG--AATTGT A-CTG-TGAG
 NA-cal   906  GTACACACCG CCCGTCGCTC T-TACCGA-- TG--AATTTC A-CTG-TGAG
 C-SIIa   895  GTACACACCG CCCGTCGCTC T-TACCGA-- TG--AATTTC C-CTG-TGAG
NA-SIIa   897  GTACACACCG CCCGTCGCTC T-TACCGA-- TG--AATTTC C-CTG-TGAG
CS-SIIa   884  GTACACACCG CCCGTCGCTC T-TACCGA-- TG--AATTTC C-CTG-TGAG
CA-SIIa   896  GTACACACCG CCCGTCGCTC T-TACCGA-- TG--AATTTC C-CTG-TGAG
CA-SIIb   886  GTACACACCG CCCGTCGCTC T-TACCGA-- TG--ACTTTC C-CTG-TGAG
NA-SIIb   886  GTACACACCG CCCGTCGCTC T-TACCGA-- TG--ACTTTC C-CTG-TGAG
  Align   662  mmmmmmmmmmm mmmmmmmmmmm m-mmmmmm-- mm--mmmmmm m-mmm-mmmm

  NA-SI   957  TTT-GCAGGA CCGA-----A CCCA------ ---------- ----------
   C-SI   957  TTT-GCAGGA CCGA-----A CCCA------ ---------- ----------
C-SI(P)   957  TTT-GCAGGA CCGA-----A CCCA------ ---------- ----------
 NA-cal   949  TTC-AATGGA CCGA-----T TTTTTCCC-- ---------- ----------
 C-SIIa   938  TTT-GAAGGA CTGA-----T GGTTG-AAAA TC-------- ----------
NA-SIIa   940  TTT-GAAGGA CTGA-----T GGTTGCAAAA TC-------- ----------
CS-SIIa   927  TTT-GAAGGA CTGA-----T GGTTGNAAAA TC-------- ----------
CA-SIIa   939  TTT-GAAGGA CTGA-----T GGTTGCAAAA TC-------- ----------
CA-SIIb   929  TTT-GAAGGA CTGG-----T GGTTGCATGA CT-------- ----------
NA-SIIb   929  TTT-GAAGGA CTGG-----T GGTTGCATGA CT-------- ----------
  Align   705  mmm-mmmmmm mmmm------ ---------- ---------- ----------

  NA-SI   975  -TTTTTGGG- TTTGGAAAT- --GCAG-TCA AA-CAGTA-C GATTTAA-AG
   C-SI   975  -TTTTTGGG- TTTGGAAAT- --GCAG-TCA AA-CAGTA-C GATTTAA-AG
C-SI(P)   975  -TTTTTGGG- TTTGGAAAT- --GCAG-TCA AA-CAGTA-C GATTTAA-AG
 NA-cal   971  ---------- TTTGGAAAT- --TTGG-TCA AA-CAGTG-A GATTTAA-AG
 C-SIIa   963  ---------- ATTGGAAAT- --TCTG-TCA AA-CAGCG-A GATTTAA-AG
NA-SIIa   966  ---------- ATTGGAAAT- --TCTG-TCA AA-CAGCG-A GATTTAA-AG
CS-SIIa   953  ---------- ATTGGAAAT- --TCTG-TCA AA-CAGCG-A GATTTAA-AG
CA-SIIa   965  ---------- ATTGGAAAT- --TCTG-TCA AA-CAGCG-A GATTTAA-AG
CA-SIIb   955  ---------- ATTGGAAAT- --TCTG-TCA AA-CAGCG-A GATTTAA-AG
NA-SIIb   955  ---------- ATTGGAAAT- --TCTG-TCA AA-CAGCG-A GATTTAA-AG
  Align   718  ---------- mmmmmimmm- --mmmm-mmm mm-mmmmm-m mmmmmmmm-mm

  NA-SI  1016  GAAAG-AGAA -G-TCGTAAC
   C-SI  1016  GAAAG-AGAA -G-TCGTAAC
C-SI(P)  1016  GAAAG-AGAA -G-TCGTAAC
 NA-cal  1004  GAAAG-AGAA -G-TCGTAAC
 C-SIIa   996  GAAAG-AGAA -G-TCGTAAC
NA-SIIa   999  GAAAG-AGAA -G-TCGTAAC
CS-SIIa   986  GAAAG-AGAA -G-TCGTAAC
CA-SIIa   998  GAAAG-AGAA -G-TCGTAAC
CA-SIIb   988  GAAAG-AGAA -G-TCGTAAC
NA-SIIb   988  GAAAG-AGAA -G-TCGTAAC
  Align   751  mmmmm-mmmm -m-mmmmmmm
```

A2.1.5. Non-spinose planktic foraminifer molecular phylogeny

```
NA-N.pac(D)   1  GCA--CCAC- AAGAACGCGT ---GGAGCA- TGTGGCTTAA --TTTGACTC
AA-N.pac(D)   1  GCA--CCAC- AAGAACGCGT ---GGAGCA- TGTGGCTTAA --TTTGACTC
NA-N.pac(S)   1  GCA--CCAC- AAGAACGCGT ---GGAGCA- TGTGGCTTAA --TTTGACTC
 C-N.dut      1  GCA--CCAC- AAGAACGCGT ---GGAGCA- TGTGGCTTAA --TTTGACTC
NA-N.dut      1  GCA--CCAC- AAGAACGCGT ---GGAGCA- TGTGGCTTAA --TTTGACTC
M-G.infl      1  GCA--CCAC- AAGAACGCGT ---GGAGCA- TGTGGCTTAA --TTTGACTC
NA-G.inf      1  GCA--CCAC- AAGAACGCGT ---GGAGCA- TGTGGCTTAA --TTTGACTC
   Align      1  mmm--mmmm- mmmmmmmmmmm ---mmmmmm- mmmmmmmmmmm --mmmmmmmmm

NA-N.pac(D)  42  AACGCGGG-- AAA--TCTTA CCAGG----- -TCCGGACAC -ACTGAG--G
AA-N.pac(D)  42  AACGCGGG-- AAA--TCTTA CCAGG----- -TCCGGACAC -ACTGAG--G
NA-N.pac(S)  42  AACGCGGG-- AAA--TCTTA CCGGG----- -TCCGGACAC -ACTGAG--G
 C-N.dut     42  AACGCGGG-- AAA--TCTTA CCGGG----- -TCCGGACAC -ACTGAG--G
NA-N.dut     42  AACGCGGG-- AAA--TCTTA CCGGG----- -TCCGGACAC -ACTGAG--G
M-G.infl     42  AACGCGGG-- AAA--TCTTA CCGGG----- -TCCGGACAC -ACTGAG--G
```

```
NA-G.inf     42  AACGCGGG-- AAA--TCTTA CCGGG----- -TCCGGACAC -ACTGAG--G
   Align     42  mmmmmmmmm-- mmm--mmmmm mmmmm----- -mmmmmmmmm -mmmmmm--m

NA-N.pac(D)  79  ATTG-ACAGG AAGTA----- --TCG--TCT TTTGAAT--T CTTTAAGGA-
AA-N.pac(D)  79  ATTG-ACAGG AAGTA----- --TCG--TCT TTTGAAT--T CTTTAAGGA-
NA-N.pac(S)  79  ATTG-ACAGG CAATA----- --TCT--CAT GTTTCATTAA CCGTTATTAA
 C-N.dut     79  ATTG-ACAGG CAATA----- --TCT--AAA TCGTTTATAA TACTTCCTAT
NA-N.dut     79  ATTG-ACAGG CAATA----- --TCT--AAA TCGTTTATAA ---TTCCTAT
M-G.infl     79  ATTG-ACAGG CAATA----- --TATTAGCA TAAAGATTCG TCTTTAGCGC
NA-G.inf     79  ATTG-ACAGG CAATA----- --TATTAGCA TAAAGACTCG TCTTTAGCGC
   Align     79  mmmm-mmmmm mmmmm----- --mm------ ---------- ----------

NA-N.pac(D) 116  -CATGTCGTT TTT--AATGA --CATCTTTT AG--ATGGAT GATTC-----
AA-N.pac(D) 116  -CATGTCGTT TTT--AATGA --CATCTTTT AG--ATGGAT GATTC-----
NA-N.pac(S) 119  CGTATCGGTT ATTTCCTTAA CATCGA---- ---------- ----------
 C-N.dut    119  TATAATACGC ATTTA----- ---------- ---------- ----------
NA-N.dut    116  TATAATACGT ATTTA----- ---------- ---------- ----------
M-G.infl    121  TAA------- ---------- ---------- ---------- ----------
NA-G.inf    121  TAA------- ---------- ---------- ---------- ----------
   Align     95  ---------- ---------- ---------- ---------- ----------

NA-N.pac(D) 154  GTG-TAAATA TGCTA-GT-- ---------- -------TCT TTCATGA---
AA-N.pac(D) 154  GTG-TAAATA TGCTA-GT-- ---------- -------TCT TTCATGA---
NA-N.pac(S) 145  GTGTTAAATA TGCTA-GT-- ---------- -------CCT TTCATGA---
 C-N.dut    134  GTGTTAAATA TGCTA-GT-- ---------- -------CCT TTCATGA---
NA-N.dut    131  GTGTTAAATA TGCTA-GT-- ---------- -------CCT TTCATGA---
M-G.infl    124  TTGTTAAATA TGCTA-GT-- ---------- -------CCT TTCATGA---
NA-G.inf    124  TTGTTAAATA TGCTA-GT-- ---------- -------CCT TTCATGA---
   Align     95  mmm-mmmmmm mmmmm-mm-- ---------- -------mmm mmmmmmmm---

NA-N.pac(D) 180  ---TTATGTG ATAG-----G TGGTG-CATG G-CCGTT-CT TAGTTCGTGG
AA-N.pac(D) 180  ---TTATGTG ATAG-----G TGGTG-CATG G-CCGTT-CT TAGTTCGTGG
NA-N.pac(S) 172  ---TTATGTG ATAG-----G TGGTG-CATG G-CCGTT-CT TAGTTCGTGG
 C-N.dut    161  ---TTATGTG ATAG-----G TGGTG-CATG G-CCGTT-CT TAGTTCGTGG
NA-N.dut    158  ---TTATGTG ATAG-----G TGGTG-CATG G-CCGTT-CT TAGTTCGTGG
M-G.infl    151  ---TTATGTG ATAG-----G TGGTG-CATG G-CCGTT-CT TAGTTCGTGG
NA-G.inf    151  ---TTATGTG ATAG-----G TGGTG-CATG G-CCGTT-CT TAGTTCGTGG
   Align    121  ---mmmmmmm mmmm-----m mmmmm-mmmm m-mmmmm-mm mmmmmmmmmm

NA-N.pac(D) 219  A-GTGATC-T GTC--TGCTT -AATTGCGTT TC-------A CTAAGGCCCC
AA-N.pac(D) 219  A-GTGATC-T GTC--TGCTT -AATTGCGTT TC-------A CTAAGGCCCC
NA-N.pac(S) 211  A-GTGATC-T GTC--TGCTT -AATTGCGTT TC-------A CTAAGGGCCC
 C-N.dut    200  A-GTGATC-T GTC--TGCTT -AATTGCGTT TC-------A CTAAGGGCCC
NA-N.dut    197  A-GTGATC-T GTC--TGCTT -AATTGCGTT TC-------A CTAAGGGCCC
M-G.infl    190  A-GTGATC-T GTC--TGCTT -AATTGCGTT TC-------A CTAAGGG-CC
NA-G.inf    190  A-GTGATC-T GTC--TGCTT -AATTGCGTT TC-------A CTAAGGGCCC
   Align    160  m-mmmmmm-m mmm--mmmmm -mmmmmmmmm mm-------m mmmmmmmmm

NA-N.pac(D) 257  A--AAGTTA- GCAAATTATC AAT--CGTTA CAGAG--T-- ------C--G
AA-N.pac(D) 257  A--AAGTTA- GCAAATTATC AAT--CGTTA CAGAG--T-- ------C--G
NA-N.pac(S) 249  A--TAAATT- -CAAGGTATG TTA--GC-TA TCGCC--GCT CTAT--G--G
 C-N.dut    238  A--TAAATT- -CAAGGTATG TTA--GC-TA TCGTT--TCT CAAT--T--G
NA-N.dut    235  A--TAAATT- -CAAGGTATG TTA--GC-TA TCGTT--TCT CAAT--T--G
M-G.infl    227  A--TAAATT- -CAAGGTATG TTA--GC-AA ATGCT--GCT CTAT--T--G
NA-G.inf    228  A--TAAATT- -CAAGGTATG TTA--GC-AA ATGCT--GCT CTAT--T--G
   Align    198  m--------- ---------- ---------- ---------- ---------m

NA-N.pac(D) 290  ACCCCT--CA CCTTTGAGT- ---------- GCGCG--TCC TAA--CTTGT
AA-N.pac(D) 290  ACCCCT--CA CCTTTGAGT- ---------- GCGCG--TCC TAA--CTTGT
NA-N.pac(S) 286  ACCCCT--TA ACCTCGGTGA A--------- GCGCG--CGT CTT--TATTT
 C-N.dut    275  ACCCCT--TG TCTTCGATAA ---------- GCGCG--TGT CTT-------
NA-N.dut    272  ACCCCT--TG TCTTCGATAA ---------- GCGCG--TGT CTT-------
M-G.infl    264  ACCCCT--AA TAGGCTTAAC TGTCTTTA-- GCGCG--TGT CTC---ACGA
NA-G.inf    265  ACCCCT--AA TAGGCTTAAC TGTCTTTA-- GCGCG--TGT CTC--TACGA
   Align    200  mmmmmm---- ---------- ---------- mmmmm----- ----------

NA-N.pac(D) 323  AGTACATGGT GTAAATGGAT TTGTT----- ---------- ----------
AA-N.pac(D) 323  AGTACATGGT GTAAATGGAT TTGTT----- ---------- ----------
NA-N.pac(S) 321  AAAGAG--TT TAAGGCAT-- ---------- --TGCGCATG CTG-------
 C-N.dut    304  TTAGAG--TT TAAA-CAT-- ---------- --TGCGCATG CTG-------
NA-N.dut    301  TTAGAG--TT TAAA-CAT-- ---------- --TGCGCATG CTG-------
M-G.infl    305  GTTC----TT TAAAGCAC-- ---------- --TGCGCATG CTG-------
NA-G.inf    307  GTTC----TT TAAAGCAC-- ---------- --TGCGCATG CTG-------
   Align    211  ---------- ---------- ---------- ---------- ----------

NA-N.pac(D) 348  ---------- ---------- ------TTGG G--TACCCA- -GAAAGCAAC
AA-N.pac(D) 348  ---------- ---------- ------TTGG G--TACCCA- -GAAAGCAAC
```

```
NA-N.pac(S)348  ---------- ---------- ------TTGG G--CTCT--- -GAAAGCAAC
 C-N.dut   330  ---------- ---------- ------TTGG G--TCCT--- -GAAAGCAAC
NA-N.dut   327  ---------- ---------- ------TTGG G--TCCT--- -GAAAGCAAC
M-G.infl   330  ---------- ---------- ------TTGG G--TCCT--- -GAAAGCAAC
NA-G.inf   332  ---------- ---------- ------TTGG G--TCCT--- -GAAAGCAAC
   Align   211  ---------- ---------- ------mmmm m--------- -mmmmmmmmmm

NA-N.pac(D)368  GAACGTGACC GC-AA----C GTCTTGT--T G--------- --------TC
AA-N.pac(D)368  GAACGTGACC GC-AA----C GTCTTGT--T G--------- --------TC
NA-N.pac(S)366  GAACGTGACC GC-AA----C GTCTTGT--T G--------- --------CC
 C-N.dut   348  GAACGTGACC GC-AA----C GTCTTAT--T G--------- --------CC
NA-N.dut   345  GAACGTGACC GC-AA----C GTCTTAT--T G--------- --------CC
M-G.infl   348  GAACGTGACC GC-AA----C GTCTTGT--T G--------- --------CC
NA-G.inf   350  GAACGTGACC GC-AA----C GTCTTGT--T G--------- --------CC
   Align   225  mmmmmmmmmm mm-mm----m mmmmmmmm--m m--------- ----------

NA-N.pac(D)394  TCTCTTTGAC AGTTATGGGT TATCCCAGTC ATGTGTTTAT ACTTTTATGT
AA-N.pac(D)394  TCTCTTTGAC AGTTATGGGT TATCCCAGTC ATGTGTTTAT ACTTTTATGT
NA-N.pac(S)392  TTAATTAAGT CGTGTTTAAA TGGTATTTGA TTACAATTTA ACCGCTTACC
 C-N.dut   374  TTTATCTTGC TATATTCTAA TTTAATTAGA AATAGCTAAC AGAGGCTAAT
NA-N.dut   371  TTTATCTTGT TATATTCTAA TTTAATTAGA AATAACTAAC AGAGGCTAAT
M-G.infl   374  TCTCTATAAT ACCTTCTTAT TTTTAATAAG AGTATTTACC TGAGGCTATT
NA-G.inf   376  TCTCTATAAT ACCTTCTTAT TTTTAATAAG AGTATTTACC TGAGGCTATT
   Align   249  ---------- ---------- ---------- ---------- ----------

NA-N.pac(D)444  GTAAATACGT ACGACACAGA GACTAGAT-- ---------- ----------
AA-N.pac(D)444  GTAAATACGT ACGACACAGA GACTAGAT-- ---------- ----------
NA-N.pac(S)442  GAGGCTATT- ---------- ---------- ---------- ----------
 C-N.dut   424  ---------- ---------- ---------- ---------- ----------
NA-N.dut   421  ---------- ---------- ---------- ---------- ----------
M-G.infl   424  ---------- ---------- ---------- ---------- ----------
NA-G.inf   426  ---------- ---------- ---------- ---------- ----------
   Align   249  ---------- ---------- ---------- ---------- ----------

NA-N.pac(D)472  ---------- ---------- ---------- ---------- ------ACC-
AA-N.pac(D)472  ---------- ---------- ---------- ---------- ------ACC-
NA-N.pac(S)451  ---------- ---------- ---------- ---------- --------A-
 C-N.dut   424  ---------- ---------- ---------- ---------- ------TTA-
NA-N.dut   421  ---------- ---------- ---------- ---------- ------TTA-
M-G.infl   424  ---------- ---------- ---------- ---------- ------TTA-
NA-G.inf   426  ---------- ---------- ---------- ---------- ------TTA-
   Align   249  ---------- ---------- ---------- ---------- ----------

NA-N.pac(D)475  --AAACTAGG CGTACCGCT- ---------- ---------- ------GTAT
AA-N.pac(D)475  --AAACTAGG CGTACCGCT- ---------- ---------- ------GTAT
NA-N.pac(S)452  --AAACTAGA CGGACCGCT- ---------- ---------- ------GTTT
 C-N.dut   427  --AAATTAGA CGGACCGCT- ---------- ---------- ------GTA-
NA-N.dut   424  --AAATTAGA CGGACCGCT- ---------- ---------- ------GTA-
M-G.infl   427  --AAACTAGA CGGACCGCT- ---------- ---------- ------GTTT
NA-G.inf   429  --AAACTAGA CGGACCGCT- ---------- ---------- ------GTTT
   Align   249  --mmmmmmmm mmmmmmmmm- ---------- ---------- ----------

NA-N.pac(D)496  CATTTCTT-- --TAAACCA- --GAGGAAGG TTGC
AA-N.pac(D)496  CATTTCTT-- --TAAACCA- --GAGGAAGG TTGC
NA-N.pac(S)473  CTTTTCT--- --TAAACCA- --GAGGAAGG TTGC
 C-N.dut   447  CTTTTCT--- --TAAACCA- --CAGGAAGG TTGC
NA-N.dut   444  CTTTTCT--- --TAAACCA- --GAGGAAGG TTGC
M-G.infl   448  CTTT-CT--- --TAAACCA- --GAGGAAGG TTGC
NA-G.inf   450  CTTT-CT--- --TAAACCA- --GAGGAAGG TTGC
   Align   266  ---------- --mmmmmmm- --mmmmmmmm mmmm
```

A2.2. Distance matrix for each within morphospecies molecular phylogeny

A2.2.1. *Globigerina bulloides* 851 bp molecular phylogeny

CS-I						
NA-IIa	0.1319					
AA-IIa	0.1319	0.0000				
AA-IIc	0.1347	0.0071	0.0071			
NA-IIb	0.1303	0.0107	0.0107	0.0155		
AA-IIb	0.1303	0.0107	0.0107	0.0155	0.0000	
CA-IId	0.1276	0.0106	0.0106	0.0130	0.0071	0.0000

A2.2.2 *Globigerina bulloides* 935 bp molecular phylogeny

CA-IId					
NA-IIb	0.0064				
AA-IIb	0.0064	0.0000			
AA-IIc	0.0196	0.0218	0.0218		
NA-IIa	0.0119	0.0119	0.0119	0.0119	
AA-IIa	0.0119	0.0119	0.0119	0.0119	0.0000

A2.2.3. *Turborotalita quinqueloba* 762 bp molecular phylogeny

CS-I				
AA-IIa	0.068			
NA-IIa	0.0681	0.0000		
NA-IIb	0.0665	0.0160	0.0160	
AA-IIc	0.0679	0.0146	0.0146	0.0013

A2.2.4. *Globigerinella* sp. 767 bp molecular phylogeny

NA-I									
C-I	0.0000								
C-I (V)	0.0000	0.0000							
NA-cal	0.0690	0.0690	0.0690						
C-IIa	0.0718	0.0718	0.0718	0.0513					
NA-IIa	0.0689	0.0689	0.0689	0.0485	0.0026				
CS-IIa	0.0704	0.0704	0.0704	0.0472	0.0052	0.0026			
CA-IIa	0.0689	0.0689	0.0689	0.0485	0.0026	0.0000	0.0026		
CA-IIb	0.0705	0.0705	0.0705	0.0543	0.0225	0.0198	0.0198	0.0198	
NA-IIb	0.0705	0.0705	0.0705	0.0543	0.0225	0.0198	0.0198	0.0198	0.0000

A2.2.5. Non-spinose planktic foraminifer 284 bp molecular phylogeny

NA-infl						
M- infl	0.0000					
NA-pac(D)	0.0326	0.0327				
AA-pac(D)	0.0326	0.0327	0.0000			
NA-pac(S)	0.0071	0.0071	0.0251	0.0251		
C-dut	0.0179	0.0180	0.0362	0.0362	0.0106	
NA-dut	0.0143	0.0143	0.0325	0.0325	0.0071	0.0035

A2 abbreviations. The genotype locations are denoted as follows. **NA** the North Atlantic, **CS** denotes the Coral Sea, **AA** the subantarctic, **CA** the Southern Californian Bight, **C** the Caribbean Sea, and **M** the Mediterranean Sea. The genotype types are denoted as I, IIa, IIb, IIc, IId. In A2.5, the morphospecies are denoted as follows: **infl** is *Globorotalia inflata*, **pac(D)** is *Neogloboquadrina pachyderma* (dextral), **pac(S)** is *Neogloboquadrina pachyderma* (sinistral), **dut** is *Neogloboquadrina dutertrei*.

Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers

Kate F. Darling, Christopher M. Wade, Iain A. Stewart, Dick Kroon, Richard Dingle & Andrew J. Leigh Brown

Reprinted from Nature, Volume 405, May 4, 2000

Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers

Kate F. Darling*†, Christopher M. Wade‡§, Iain A. Stewart*, Dick Kroon*, Richard Dingle‖§ & Andrew J. Leigh Brown†

* *Department of Geology and Geophysics, University of Edinburgh, Edinburgh EH9 3JW, UK*
† *Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK*
‡ *Institute of Genetics, University of Nottingham, Nottingham NG7 2UH, UK*
‖ *British Antarctic Survey, Madingley Road, Cambridge CB3 0ET, UK*

Bipolarity, the presence of a species in the high latitudes separated by a gap in distribution across the tropics, is a well-known pattern of global species distribution. But the question of whether bipolar species have evolved independently at the poles since the establishment of the cold-water provinces 16–8 million years ago, or if genes have been transferred across the tropics since that time, has not been addressed. Here we examine genetic variation in the small subunit ribosomal RNA gene of three bipolar planktonic foraminiferal morphospecies. We identify at least one identical genotype in all three morphospecies in both the Arctic and Antarctic subpolar provinces, indicating that trans-tropical gene flow must have occurred. Our genetic analysis also reveals that foraminiferal morphospecies can consist of a complex of genetic types. Such occurrences of genetically distinct populations within one morphospecies may affect the use of planktonic foraminifers as a palaeoceanographic proxy for climate change and necessitate a reassessment of the species concept for the group.

Patterns of bipolar (anti-tropical) distributions are found in a diverse range of terrestrial and marine groups[1,2]. Bipolar organisms were first observed by biologists during the mid-nineteenth century[3–5], and are now known to be recurrent phenomena through considerable spans of geological time[1]. Within the marine environment, bipolar species are observed in many planktonic groups. It remains unclear whether they are isolated within their respective high-latitude water masses, or whether trans-tropical transit occurs; such transit would allow intermixing and consequent gene flow between the populations from the northern and the southern hemispheres. Morphology alone can provide few clues, as morphological identity may be due to convergent or parallel evolution in similar environments. Molecular data enable us to investigate this issue directly. If high-latitude populations of bipolar taxa were isolated from one another, different mutations would be expected to accumulate over time, leading to genetic divergence between the two populations. Conversely, if trans-tropical mixing occurs resulting in genetic exchange, then the two populations would be expected to be genetically homogeneous.

Advances in planktonic foraminiferal molecular genetic analysis[6,7] provide a tool with which to investigate genetic interchange in marine planktonic organisms exhibiting a bipolar distribution[8]. The planktonic foraminifera have an outstanding fossil record, and their calcitic shells (tests) form one of the most widely used microfossil assemblages for the reconstruction of past oceanic environments. They are globally distributed, and their component taxa are found within distinct faunal provinces that broadly correspond to the main hydrographic features of the ocean[8] (Fig. 1). Planktonic foraminiferal nucleotide sequence data now provide a new dimension for the investigation of oceanic gene flow[9] which will enable us to answer the key question of whether genetic exchange occurs between high-latitude populations.

§ Present addresses: Department of Zoology, The Natural History Museum, London SW7 5BD, UK (C.M.W.); Geological Institute, University of Copenhagen, Oster Voldgade 10, DK-1350 Copenhagen K, Denmark (R.D.).

Molecular evolution of planktonic foraminifera

Small subunit ribosomal DNA (SSU rDNA) sequences are highly variable in planktonic foraminifera, and phylogenetic analysis has revealed that many morphologically defined species (morphospecies) of planktonic foraminifera in fact represent complexes of different and often highly divergent genetic types (genotypes)[9]. Some of these genotypes are now considered to be cryptic sibling species[6,10,11], a commonly observed phenomenon amongst marine taxa[12]. We have analysed SSU rDNA sequences of cool-water representatives of three planktonic foraminiferal morphospecies—*Globigerina bulloides* d'Orbigny, *Turborotalita quinqueloba* (Natland) and *Neogloboquadrina pachyderma* (Ehrenberg)—that exhibit a predominantly disjunct, bipolar distribution[8]. Arctic specimens were collected from the south of Iceland to the southeast Greenland margin along two transects, and Antarctic specimens were collected from eight sites along a transect between the Falkland Islands and the Antarctic Peninsula (Fig. 1; Methods).

A foraminiferal SSU rDNA phylogeny incorporating these species is shown in Fig. 2. The phylogeny is based on a restricted data set comprising only those nucleotide positions that could be aligned across all 22 genera (see Fig. 2 legend for details). The placement of *G. bulloides* and *T. quinqueloba* within the spinose cluster in the molecular phylogeny is consistent with the evolutionary relationships inferred from palaeontological data[13,14]. The location of right- and left-coiling *N. pachyderma* within the benthic region of the tree is, however, inconsistent with palaeontological data, providing further evidence that the extant planktonic foraminifera are not monophyletic in origin[6,7].

Each of the *G. bulloides*, *T. quinqueloba* and *N. pachyderma* morphospecies consist of a complex of distinct genotypes (Fig. 2 and Table 1; individual genotypes have identical SSU rDNA sequences). Both left- and right-coiling *N. pachyderma* genotypes cluster with *Neogloboquadrina dutertrei* which forms the warm-water member of the morphological complex[15] (Fig. 2). However, the relationships observed within this cluster are surprising. Unexpectedly, *N. pachyderma* left-coiling genotypes cluster together with

N. dutertrei, which is right-coiling, and away from the *N. pachyderma* right-coiling genotype. Thus the genetic data we report here confirm that coiling direction in *N. pachyderma* is associated with genetic divergence[16] and not with temperature during development.

Bipolar SSU rDNA genotypes in foraminiferal morphospecies

When considering identity between genotypes, we utilize not only the conserved regions used for the phylogenetic analysis but also the remaining ~500 nucleotide sites from the highly informative variable regions (Fig. 3). Such comparisons of the entire SSU rDNA fragment revealed that all three morphospecies examined had at least one identical SSU rRNA genotype in both the Arctic and Antarctic subpolar provinces (Figs 2 and 3, and Table 1). There were two separate bipolar genotypes (types IIa and IIb) in *G. bulloides*, one bipolar genotype (type IIa) in *T. quinqueloba* and one bipolar genotype (type R) in *N. pachyderma* right-coiling.

The *G. bulloides* genotypes, types IIa and IIb, differed from one another at 36 of 980 nucleotide sites (variable and conserved regions; Fig. 3 and data not shown). There were also two other closely related genotypes in the *G. bulloides* cluster. Type IIc was found only in the subantarctic province, and differed from the type IIa bipolar genotype at 31 sites and from the type IIb bipolar genotype at 49 sites (Fig. 3 and data not shown). Type IId was found in the transitional zone[9],and differed from the type IIa bipolar genotype at 43 sites and from the type IIb bipolar genotype at 18 sites (Fig. 3 and data not shown). Two further genotypes were also found in *T. quinqueloba* that were closely related to the bipolar type IIa: type IIb in the subarctic and type IIc in the subantarctic provinces. Type IIb differs from type IIc in only 8 of 1,168 nucleotide sites (variable and conserved) and these differ from the bipolar Type IIa genotype at 83 and 82 nucleotide sites, respectively. The Antarctic type III and type IV genotypes of *N. pachyderma* (left-coiling) differ from one another by as many as 145 nucleotide sites (Fig. 3 and data not shown). Thus, where detailed comparisons may be made across both conserved and variable regions, there is a continuum of genetic divergence in the cool-water 'genotype clusters' extending from identical bipolar genotypes through very closely related variants to others which differ at more than 10% of sites in the region sequenced.

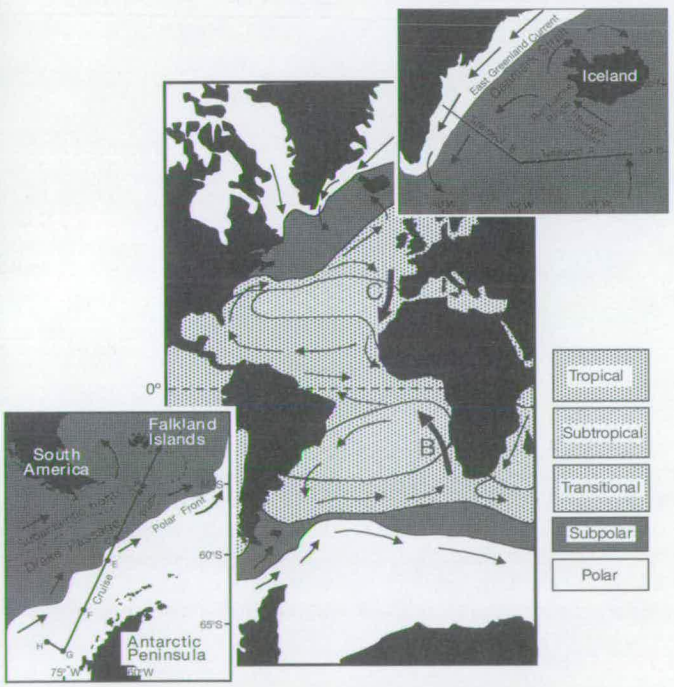


Figure 1 Sampling localities and distribution of the high-latitude morphospecies. The pictorial map (main figure) shows the five zoogeographical planktonic foraminiferal faunal provinces[8]— polar, subpolar, transitional, subtropical and tropical—of the North and South Atlantic Ocean. The three planktonic foraminiferal morphospecies *G. bulloides*, *T. quinqueloba* and *N. pachyderma* (right- and left-coiling) used in this study are found predominantly in the polar, subpolar and transitional provinces, and consequently exhibit a bipolar (anti-tropical) distribution[8]. The cool-water eastern boundary cold currents, the Canary and Benguela (C and B, large arrows), strengthen seasonally and flow equatorwards, bringing cool water and transitional morphospecies into the subtropical province where they bloom in the transient cool seasonal upwelling systems of the subtropical/tropical eastern Atlantic. However, there is always a clear belt of warm tropical water moving seasonally north and south as the seasons build and decay. An outline of the present-day Atlantic Ocean surface circulation patterns (small arrows) are shown. The expanded maps (insets) highlight the sampling transect, sample sites (see Methods) and surface currents within the Arctic and Antarctic subpolar/polar provinces.

Genetic exchange between high-latitude populations

Foraminiferal sequence data have allowed us to address the fundamental question of whether gene flow occurs between bipolar-distributed planktonic marine organisms. The presence of identical foraminiferal sequences in the high latitudes of each hemisphere, coupled with their extensive fossil record, has allowed us to establish that trans-tropical genetic exchange has been recent. The well-calibrated foraminiferal biostratigraphic record permits the estimation of the expected level of divergence between two sequences over time in the absence of gene flow. Rates of evolution for foraminiferal SSU rRNA genes have been estimated by both our group[9] and others[17,18]. These estimates have given approximate rates of substitution of $(2.6–4.6) \times 10^{-9}$ substitutions per site per year for the spinose planktonic group[9,18]. Taking an average, we would expect 1.82 substitutions per million years in the conserved 505-base-pair (505-b.p.) region alone, or 29–15 substitutions since the cold-water provinces were established 16–8 million years ago in the Middle to Late Miocene epoch[19].

In addition to the conserved regions, intra-specific comparisons can be made on the variable regions that cannot be utilized for between-morphospecies analysis (Fig. 2). On the basis of our data, the rate of substitution over the variable regions is estimated to be at least 10 times greater than the rate for the conserved regions: thus we expect between 160 and 320 substitutions over the entire sequence since the establishment of the polar provinces. These calculations establish that a substantial degree of divergence would be expected if sequences were isolated for between 8 and 16 million years, yet we have identified examples of complete sequence identity between the high-latitude morphospecies. This strongly suggests that genetic exchange has occurred recently relative to the establishment of the cold-water provinces. This conclusion is reinforced by the fact that it was observed independently in three separate lineages.

Trans-tropical gene flow

Genetic exchange between the high latitudes can only be explained by trans-equatorial transit. The tropical ocean is, however, a formidable barrier for cool-water genotypes to cross (Fig. 1). Foraminiferans do not encyst—a process that provides resistance to inhospitable environments in many other planktonic groups. It is therefore unlikely that the subpolar genotypes would tolerate the high surface water temperatures of the western tropical regions. Yet for complete bipolar genetic mixing to occur, foraminifers crossing the tropics must also return to the high latitudes within the warm tropical ocean gyral surface currents (Fig. 1).

How then can gene flow occur between the cooler-water populations across the tropical Atlantic? The present-day oceanographic setting provides a perspective in which to investigate this issue. The eastern boundaries of the subtropical Atlantic Ocean off west Africa are associated with cool boundary currents (shown as B and C in Fig. 1) which act as corridors for the introduction of cool-water

genotypes into the cool seasonal upwelling zones of the subtropical region, where they may bloom in significant numbers. The upwelling cells could then provide a 'stepping stone' for the transit of the cool-water genotypes into the permanent equatorial upwelling zone (which is 2–9 °C cooler than surrounding surface water). Yet genetic exchange between populations within these waters alone cannot account for the genetic homogeneity observed between the subpolar populations. For bipolar genetic exchange to occur, the foraminifera would have to negotiate the high tropical sea surface temperatures of the western equatorial region into which they are passively carried. This region is most unlikely to be conducive to the survival of the cool-water genotypes, either as a direct result of temperature

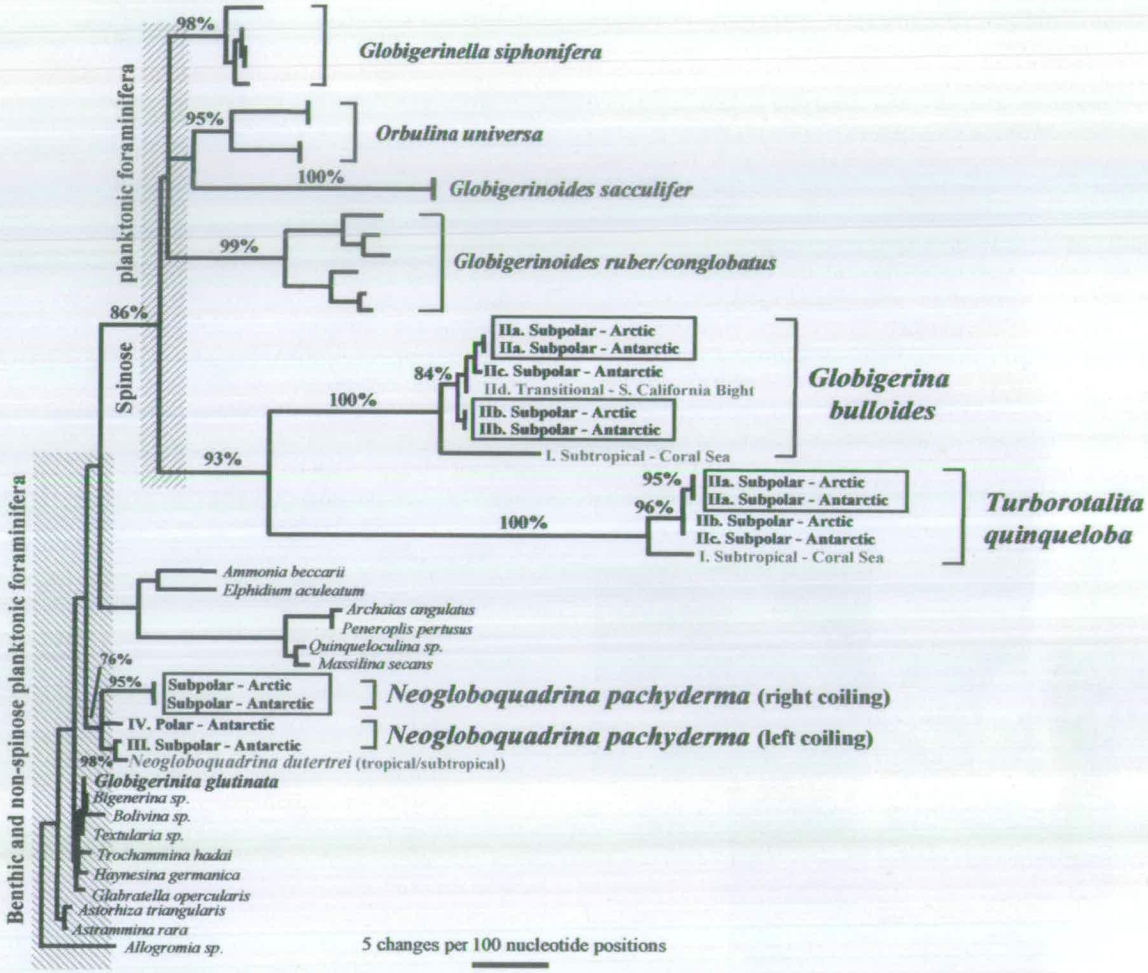


Figure 2 Foraminiferal SSU rDNA phylogeny, highlighting the evolutionary placement of the planktonic foraminiferal high-latitude morphospecies. The phylogeny shows the genotypic variants identified for *G. bulloides* d'Orbigny ($n = 55$), *T. quinqueloba* (Natland) ($n = 24$) and *N. pachyderma* (Ehrenberg) right-coiling ($n = 26$) and left-coiling ($n = 31$). *N. pachyderma* falls within the benthic and non-spinose planktonic foraminiferal region of the tree, and *G. bulloides* and *T. quinqueloba* fall within the spinose planktonic foraminiferal group (shaded boxes). Genotypes of *G. bulloides* and *T. quinqueloba* isolated from the transitional and subtropical regions are shown in grey. *N. dutertrei*, the warm-water member of the *Neogloboquadrina* morphological complex is also shown in grey. High-latitude genotypes of *G. bulloides*, *T. quinqueloba* and *N. pachyderma* (right- and left-coiling) are shown in black. Genotypes which are identical in both the Arctic and Antarctic (bipolar genotypes) are boxed. Bootstrap values (expressed as a percentage) are shown for the main branches when supported in >70% of replicates. Interrelationships within the transitional and subpolar *G. bulloides* cluster are unresolved. The number of specimens obtained for each genotype and their location (Fig. 1) are shown in Table 1.

Table 1 Number and location of specimens obtained for each genotype.

Morphospecies	Location	Genotype	Specimens	Sample Sites
G. bulloides	Subpolar–Arctic ($n = 44$)	IIa	32	Subarctic transect B
		IIb	12	Subarctic transect A
	Subpolar–Antarctic ($n = 11$)	IIa	5	Site A + B + C + D
		IIb	4	Site D
		IIc	2	Site A
T. quinqueloba	Subpolar–Arctic ($n = 17$)	IIa	8	Subarctic transect A + B
		IIb	9	Subarctic transect A + B
	Subpolar–Antarctic ($n = 7$)	IIa	5	Site C + D
		IIc	2	Site A
N. pachyderma	Subpolar–Arctic ($n = 24$)	Right	24	Subarctic transect A + B
	Subpolar–Antarctic ($n = 14$)	Right	2	Site A
		Left III	12	Site B + C + D
	Polar–Antarctic ($n = 19$)	Left III	3	Site E + F
		Left IV	16	Site E + F + G + H

Locations identified in Fig. 1.

or as a consequence of other physical/biological requirements being limiting within the ecosystem. It is possible that transit occurs through these regions by tropical submergence into cooler levels of the thermocline. The route and timing of transit is likely to be morphospecies-specific due to differences in their capacity to survive transit across the tropics. Whether gene flow is uni-directional or bi-directional remains to be determined.

Trans-equatorial transit and subsequent genetic exchange is likely to increase substantially during cooling periods associated with glacial cycling in the Quaternary period (the past 1.8 million years), and indeed the low-latitude sedimentary record shows the frequent occurrence of planktonic foraminiferal subpolar assemblages within the equatorial zone during these cooling cycles[20]. Revised estimates of tropical cooling during the Last Glacial Maximum (18,000 ^{14}C years before present)[21–24] indicate a substantial drop in temperature of 3–5 °C in the tropical Atlantic, accompanied by an equatorward extension and intensification of the cool-water boundary currents[25–27] (B and C in Fig. 1). This would have increased the potential for genetic exchange between the bipolar populations. It is, however, possible that genetic exchange could be continuous and may be occurring at present. The available data are unable to resolve timescales of this magnitude, and the issue can only be addressed by sampling living planktonic foraminifers in the subtropical/tropical regions to establish the distribution pattern, proportion and seasonality of cool-water genotypes in the present day.

Speciation in planktonic foraminifers

Whereas molecular studies have shown that many planktonic foraminiferal morphospecies comprise more than one genetically distinct population which may warrant 'cryptic' species status[6,9,10,11], we have also shown that planktonic foraminiferal populations intermix on a global scale (Arctic–Antarctic, this study; Pacific–Atlantic, refs 6 and 9). This makes any simple speciation model based on geographical isolation (allopatric speciation) difficult to sustain. Yet allopatric isolation may occur in the marine environment, and there are many examples of population subdivision in marine species despite a high dispersal potential[28]. Subdivision may result from the sporadic and discontinuous seeding of seasonal upwelling cells or alternating ocean circulation patterns associated with past climate change. Such isolation could give rise to genetically distinct populations.

In order for speciation to occur, genetic divergence must be accompanied by reproductive isolation[29]. In marine organisms with broadcast release of gametes such as the planktonic foraminifera, this may be biological, involving, for example, synchrony of gametogenesis[10,28,30]. Alternatively, isolation may be molecular and there is evidence that this may be achieved through high rates of evolution in proteins involved in gamete recognition[31,32]. Recent developments in population genetic theory of sympatric speciation[33,34] suggest further possibilities for speciation in the marine planktonic environment without the necessary initial requirement for allopatry. We conclude that there are several currently recognized mechanisms that could be involved in the speciation of planktonic foraminifera, but the main problem with interpretation of these processes is lack of knowledge of the structuring of their environment in either biotic or abiotic terms. The discovery of genotype complexes in the planktonic foraminifera does, however, necessitate the urgent reassessment of species concepts for the group.

Distribution of planktonic foraminiferal cool-water genotypes

The high-latitude genotypes are not ubiquitous in the surface waters throughout the cool-water provinces. For example, in the Arctic subpolar region, *G. bulloides* type IIa was largely confined to transect B and type IIb to transect A (Fig. 1, Table 1). Given that sampling density was relatively low in the Antarctic subpolar region (due to adverse weather conditions during sampling), *G. bulloides* type IIa was distributed from sites A–D—from the Falkland Islands to the polar front—but type IIb was found only at the polar front (site D, Fig. 1, Table 1). Similarly, *N. pachyderma* left-coiling type III was mostly found between the Falkland Islands and the polar front (sites B–D, Table 1) and type IV was isolated in the true polar waters (sites E–H, Table 1). Although we have no direct evidence at present, such distribution patterns may indicate that individual genotypes are adapted to specific hydrographic or trophic environments; we note that de Vargas *et al.*[11] found a close correlation between the distribution of sibling species and trophic regimes in the spinose species *Orbulina universa*.

Implications for palaeoceanography and palaeoclimate

The morphological, chemical and stable-isotope differences associated with planktonic foraminiferal calcitic shells are used extensively by palaeoceanographers for climate reconstruction. For such studies, the assumption is made that each morphospecies represents a genetically continuous species with a single environmental/habitat preference. If this is not the case—as indicated by this study and others[6,9,10,11]—stable-isotope and geochemical analyses of planktonic foraminiferal shells, and census-based transfer-function techniques derived from such pooled data, must include significant

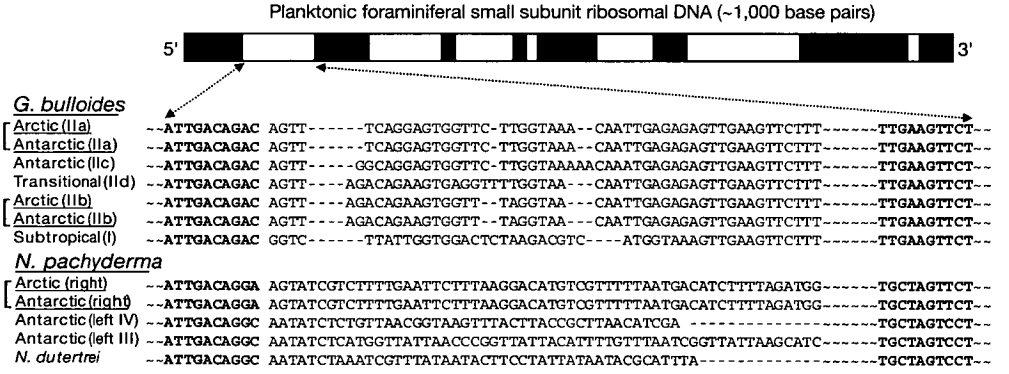


Figure 3 Sequence variability in the SSU rRNA gene. Shown is a schematic representation of the ~1,000-b.p. region of the SSU rRNA gene sequenced for the planktonic foraminifers. Conserved regions alignable across all foraminiferal taxa (505 b.p.) and used in phylogenetic tree construction (see Fig. 2) are highlighted in black. Variable regions, not used in phylogenetic tree construction because they could not be aligned across all taxa, are highlighted in white. Examples of the sequence variation observed in one of the variable regions are given for *G. bulloides* and *N. pachyderma/dutertrei*. Considerable variability is observed within the variable region, even among closely related cool-water genotypes. The bipolar genotypes (marked with a bracket and underlined) are identical across the entire sequenced region. (~ denotes sequences not displayed; – denotes gaps introduced in aligning sequences).

noise, if not error. If genetic differences can be correlated with specific environment and habitat preferences, and the genotypes differentiated in the fossil record, a new level of precision for climate modelling could be achieved. □

Methods

Sampling localities

Arctic subpolar specimens of *G. bulloides*, *T. quinqueloba* and *N. pachyderma* right-coiling were collected on board RV *Professor Logachev* along two transects (A and B, Fig. 1). Transect A ran from a point 59° 58′ N, 11° 34′ W (south of Iceland) to 58° 56′ N, 30° 24′ W (above the Reykjanes ridge). Transect B ran from the Reykjanes ridge to the southeast Greenland margin at 63° 48′ N, 40° 20′ W across the Denmark Strait. Specimens were obtained by pumping continually from 4.5 m depth through a plankton net suspended on deck. Antarctic subpolar specimens of *G. bulloides*, *T. quinqueloba* and *N. pachyderma* left- and right-coiling were collected from eight sites (A–H) between the Falkland Islands (53° 21′ S, 58° 20′ W) and west of the Antarctic Peninsula (65° 36′ S, 77° 39′ W) (inset map, Fig. 1). Specimens were obtained by pumping from 6 m depth through a 63-μm filter on board RRS *James Clark Ross* (JR19). Transitional specimens of *G. bulloides* were collected ~2 km NNE of the Santa Catalina Marine Science Centre, Santa Catalina Island, California (33° N, 118° W), and subtropical specimens of *G. bulloides* and *T. quinqueloba* were collected off the Great Barrier Reef, Australia (0.8 nautical miles due east of Ribbon Reef 10).

Isolation and sequencing of SSU genes

DNA extraction, amplification by polymerase chain reaction and direct automated sequencing of an ~1,000-b.p. region of the terminal 3′ end of the foraminiferal SSU rRNA gene was as described previously[6,9,35]. The SSU rRNA genes form a large multigene family in all organisms where data are available. Within the gene family, as mutations accumulate in individual copies, a level of homogeneity is maintained between copies by the process of concerted evolution. All data presented here have been derived from a consensus sequence amplified from the gene family of a single individual using a direct sequencing approach. This has the advantage of eliminating the risk of sampling non-orthologous copies, as different copies would be detected as ambiguities in the consensus sequence. Using this approach we have detected no evidence of ambiguity in any of the genotypes presented here, apart from very minor variability between copies within individuals of left-coiling *N. pachyderma*.

Sequence analysis

Partial SSU rDNA sequences were aligned manually within version 2.2 of the Genetic Data Environment (GDE) package[36]. 505 unambiguously aligned nucleotide sites were employed in phylogeny reconstruction. The regions included are indicated schematically in Fig. 3 ("conserved regions"). A neighbour-joining[37] phylogeny was generated using γ rate corrected F84 distances within PAUP* version 4.0d64 (kindly provided by D. L. Swofford[38]). α values for rate correction were estimated within PAUP*. Bootstrap resampling[39] (2,000 replicates) was employed to assign support to the neighbour-joining tree. Cool-water planktonic foraminiferal taxa are placed within a background of additional planktonic[9] and benthic[17] foraminifers. The tree was rooted on *Allogromia*, a membranous-walled benthic foraminifer which is thought to have diverged early in the history of the clade[40].

Received 19 April 1999; accepted 24 February 2000.

1. Crame, J. A. Bipolar molluscs and their evolutionary implications. *J. Biogeogr.* **20**, 145–161 (1993).
2. Lindberg, D. R. Marine biotic interchange between the northern and southern hemispheres. *Paleobiology* **17**, 308–324 (1991).
3. Darwin, C. *The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (John Murray, London, 1859).
4. Hooker, J. D. *The Botany. The Antarctic Voyage of H. M. Discovery Ships Erebus and Terror, in the Years 1839-1843* (Reeve Brothers, London, 1847).
5. Ross, J. C. *A Voyage of Discovery and Research in the Southern and Antarctic Regions, during the Years 1839-43* (John Murray, London, 1847).
6. Darling, K. F., Wade, C. M. Kroon, D. & Leigh Brown, A. J. Planktic foraminiferal molecular evolution and their polyphyletic origins from benthic taxa. *Mar. Micropaleontol.* **30**, 251–266 (1997).
7. de Vargas, C., Zaninetti, L., Hilbrecht, H. & Pawlowski, J. Phylogeny and rates of molecular evolution of planktonic foraminifera: SSU rDNA sequences compared to the fossil record. *J. Mol. Evol.* **45**, 285–294 (1997).
8. Bé, A. W. H. in *Oceanic Micropaleontology* (ed. Ramsey, A. T. S.) 1–100 (Academic, London, 1977).
9. Darling, K. F., Wade, C. M., Kroon, D., Leigh Brown, A. J. & Bijma, J. The diversity and distribution of modern planktic foraminiferal small subunit ribosomal RNA genotypes and their potential as tracers of present and past ocean circulations. *Paleoceanography* **14**, 3–12 (1999).
10. Huber, B. T., Bijma, J. & Darling, K. F. Cryptic speciation in the living planktonic foraminifer *Globigerinella siphonifera* (d'Orbigny). *Paleobiology* **23**, 33–62 (1997).
11. de Vargas, C., Norris, R., Zaninetti, L., Gibb, S. W. & Pawlowski, J. Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. *Proc. Natl Acad. Sci. USA* **96**, 2864–2868 (1999).
12. Knowlton, N. Sibling species in the sea. *Annu. Rev. Ecol. Systematics* **24**, 189–216 (1993).
13. Kennett, J. P. & Srinivasan, M. S. *Neogene Planktonic Foraminifera, a Phylogenetic Atlas* (Hutchinson and Ross, Stroudsburg, Pennsylvania, 1983).
14. Spezzaferri, S. Planktonic foraminiferal biostratigraphy of the Oligocene and lower Miocene in the oceanic record. An overview. *Palaeontogr. Italica* **81**, 1–187 (1994).
15. Hilbrecht, H. Morphologic gradation and ecology in *Neogloboquadrina pachyderma* and *N. dutertrei* (planktic foraminifera) from core top sediments. *Mar. Micropaleontol.* **31**, 31–43 (1997).
16. Brummer, G. A. & Kroon, D. in *Planktonic Foraminifers as Tracers of Ocean-climate History* (eds Brummer, G. A. & Kroon, D.) 293–298 (Free University, Amsterdam, 1988).
17. Pawlowski, J. *et al.* Extreme differences in rates of molecular evolution of foraminifera revealed by comparison of ribosomal DNA sequences and the fossil record. *Mol. Biol. Evol.* **14**, 498–505 (1997).
18. de Vargas, C. & Pawlowski, J. Molecular versus taxonomic rates of evolution in planktonic foraminifera. *Mol. Phyl. Evol.* **9**, 463–469 (1998).
19. Kennett, J. P., Keller, G. & Srinivasan, M. S. Miocene planktonic foraminiferal biogeography and paleoceanographic development of the Indo-Pacific region. *Geol. Soc. Am.* **163**, 197–236 (1985).
20. McIntyre, A., Ruddiman, W. F., Karlin, K. & Mix, A. C. Surface water response of the equatorial Atlantic Ocean to orbital forcing. *Paleoceanography* **4**, 19–55 (1989).
21. Guilderson, T. P., Fairbanks, R. G. & Rubenstone, J. L. Tropical temperature variations since 20,000 years ago: modulating interhemispheric climate change. *Science* **263**, 663–665 (1994).
22. Stute, M. *et al.* Cooling of tropical Brazil (5°C) during the last glacial maximum. *Science* **269**, 379–383 (1995).
23. Schrag, D. P., Hampt, G. & Murray, D. Pore fluid constraints on the temperature and oxygen isotopic composition of the glacial ocean. *Science* **272**, 1930–1932 (1996).
24. Wolff, T., Mulitza, S., Arz, H., Pätzold, F. & Wefer, G. Oxygen isotopes versus CLIMAP (18ka) temperatures: a comparison from the tropical Atlantic. *Geology* **26**, 675–678 (1998).
25. CLIMAP Project Members. Seasonal reconstructions of the Earth's surface at the last glacial maximum. *Geol. Soc. Am. Map Chart Ser.* **MC-36**, 1–18 (1981).
26. Mix, A. C. & Morey, A. E. in *The South Atlantic: Present and Past Circulation* (eds Wefer, G., Berger, W. H., Siedler, G. & Webb, D. J.) 503–525 (Springer, Berlin, 1996).
27. Zhao, M., Beveridge, N. A. S., Shackleton, N. J., Sarnthein, M. & Eglinton, G. Molecular stratigraphy of cores off northwest Africa: Sea surface temperature history over the last 80 ka. *Paleoceanography* **10**, 661–675 (1995).
28. Palumbi, S. R. Genetic divergence, reproductive isolation, and marine speciation. *Annu. Rev. Ecol. Systematics* **25**, 547–572 (1994).
29. Coyne, J. A. & Orr, H. A. The evolutionary genetics of speciation. *Phil. Trans. R. Soc. Lond. B* **353**, 287–305 (1998).
30. Bijma, J., Erez, J. & Hemleben, C. Lunar and semi-lunar reproductive cycles in some spinose planktonic foraminifers. *J. Foram. Res.* **20**, 117–127 (1990).
31. Metz, E. C. & Palumbi, S. R. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Mol. Biol. Evol.* **13**, 397–406 (1996).
32. Biermann, C. H. The molecular evolution of sperm bindin in six species of sea urchins (Echinoida: Strongylocentrotidae). *Mol. Biol. Evol.* **15**, 1761–1771 (1998).
33. Dieckmann, U. & Doebeli, M. On the origin of species by sympatric speciation. *Nature* **400**, 354–357 (1999).
34. Tregenza, T. & Butlin, R. K. Speciation without isolation. *Nature* **400**, 311–312 (1999).
35. Darling, K. F., Kroon, D., Wade, C. M. & Leigh Brown, A. J. Molecular evolution of planktic foraminifera. *J. Foram. Res.* **26**, 324–330 (1996).
36. Smith, S. W., Overbeek, R., Woese, C. R., Gilbert, W. & Gillevet, P. M. The genetic data environment, an expandable GUI for multiple sequence analysis. *Comput. Appl. Biosci.* **10**, 671–675 (1994).
37. Saitou, N. & Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
38. Swofford, D. L. PAUP*. Phylogenetic Analysis using Parsimony (*and Other Methods) V. 4 (Sinauer Associates, Sunderland, Massachusetts, 1998).
39. Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791 (1985).
40. Tappan, H. & Loeblich, A. R. Foraminiferal evolution, diversification and extinction. *J. Paleontol.* **62**, 695–714 (1988).

Acknowledgements

Acknowledgements We thank C. Pudsey and T. Paramour for their assistance during the 1997 Antarctic cruise on RRS *James Clark Ross* (British Antarctic Survey), and S. Troelstra and colleagues for their support during the 1997 Denmark Strait expedition on RV *Professor Logachev* (Netherlands Foundation for Scientific Research and the Geological Survey of Denmark and Greenland). We also thank A. Davison, S. Goodacre, J. Brookfield, P. Sharp and B. Clarke for discussions and comments on the manuscript. This work was supported by the Leverhulme Trust, the Carnegie Trust for the Universities of Scotland and the NERC.

Correspondence and requests for materials should be addressed to K.F.D. (e-mail: Kate.Darling@ed.ac.uk). The sequences determined for this study are available in GenBank under the accession numbers AF250105–AF250120.